

Counting in Graph Covers: A Combinatorial Characterization of the Bethe Entropy Function

Pascal O. Vontobel

Abstract—We present a *combinatorial* characterization of the Bethe entropy function of a factor graph, such a characterization being in contrast to the original, *analytical*, definition of this function. We achieve this combinatorial characterization by counting valid configurations in finite graph covers of the factor graph.

Analogously, we give a *combinatorial* characterization of the Bethe partition function, whose original definition was also of an *analytical* nature. As we point out, our approach has similarities to the replica method, but also stark differences.

The above findings are a natural backdrop for introducing a decoder for graph-based codes that we will call symbolwise graph-cover decoding, a decoder that extends our earlier work on blockwise graph-cover decoding. Both graph-cover decoders are theoretical tools that help towards a better understanding of message-passing iterative decoding, namely blockwise graph-cover decoding links max-product (min-sum) algorithm decoding with linear programming decoding, and symbolwise graph-cover decoding links sum-product algorithm decoding with Bethe free energy function minimization at temperature one.

In contrast to the Gibbs entropy function, which is a concave function, the Bethe entropy function is in general not concave everywhere. In particular, we show that every code picked from an ensemble of regular low-density parity-check codes with minimum Hamming distance growing (with high probability) linearly with the block length has a Bethe entropy function that is convex in certain regions of its domain.

Index Terms—Bethe approximation, Bethe entropy, Bethe partition function, graph cover, graph-cover decoding, message-passing algorithm, method of types, linear programming decoding, pseudo-marginal vector, sum-product algorithm.

I. INTRODUCTION

WHAT IS THE meaning of the pseudo-marginal functions that are computed by the sum-product algorithm, especially at a fixed point of the sum-product algorithm? This question stood at the beginning of our investigations. For factor graphs *without* cycles the answer is clear, and was stated succinctly already by Wiberg *et al.* [1], [2]: the pseudo-marginal functions at a fixed point of the sum-product algorithm (SPA) are the correct marginal functions of the global function that is represented by the factor graph. Note that here and hereafter we assume that SPA messages are updated according to the so-called “flooding message update

schedule.” For factor graphs of finite size and without cycles this implies that the SPA reaches a fixed point after a finite number of iterations [1], [2].

However, in the case of factor graphs *with* cycles, the answer is a priori not so clear, even for fixed points of the SPA.¹ Of course, one can express the SPA-based pseudo-marginal functions as marginal functions of the global functions of computation-tree factor graphs, the latter being unwrapped versions of the factor graph under consideration [1], [2]. However, the analysis of these objects has so far proven to be rather difficult, a main reason for this being that the computation trees, along with the global functions represented by them, change with the iteration number.

A. A Combinatorial Characterization of the Bethe Entropy Function in Terms of Finite Graph Covers

Towards making progress on the above-mentioned question, this paper studies the Bethe free energy function of a factor graph, a function that was introduced by Yedidia, Freeman, and Weiss [3] and whose importance stems from a very well-known theorem in [3] which states that fixed points of the SPA correspond to stationary points of the Bethe free energy function. Consequently, it is clearly desirable to obtain a better understanding of this function by characterizing it from different perspectives.

Recall that the Bethe free energy function F_B is defined to be

$$F_B(\beta) \triangleq U_B(\beta) - T \cdot H_B(\beta),$$

where β is a (locally consistent) pseudo-marginal vector, U_B is the Bethe average energy function, H_B is the Bethe entropy function, and $T \geq 0$ is the temperature. (All the mathematical terms appearing in this introduction will be suitably defined in later sections.) Both U_B and H_B contribute significantly towards the shape of F_B . However, the curvature of F_B is exclusively determined by the curvature of H_B ; this is a consequence of the fact that U_B is a linear function of its argument. Therefore, characterizing the function F_B is nearly tantamount to characterizing the function H_B .

In this paper we offer a combinatorial characterization of the Bethe entropy function H_B in terms of finite graph covers of the factor graph under consideration. Recall that in earlier

Submitted to IEEE Transactions on Information Theory, November 20, 2010; revised on September 22, 2012; date of current version, October 9, 2012. Some of the material in this paper was previously presented at the 46th Annual Allerton Conference on Communications, Control, and Computing, Monticello, IL, USA, Sep. 23–26, 2008, at the 2009 Information Theory Workshop, Taormina, Italy, Oct. 10–13, 2009, and at the 48th Annual Allerton Conference on Communications, Control, and Computing, Monticello, IL, USA, Sep. 29–Oct. 1, 2010.

P. O. Vontobel is with Hewlett-Packard Laboratories, 1501 Page Mill Road, Palo Alto, CA 94304, USA (e-mail: pascal.vontobel@ieee.org).

¹ Here and in the following we assume that the local functions of the factor graph are non-negative. Moreover, we assume that we are only interested in SPA-based pseudo-marginal functions that are normalized, *i.e.*, pseudo-marginal functions that sum to one. Therefore, without loss of generality, we may assume that at every iteration the SPA messages are normalized, *i.e.*, that they sum to one. With this, SPA fixed points are well defined, also for factor graphs with cycles.

work [4], [5] we showed that every valid configuration in some finite graph cover maps down to some pseudo-marginal vector with rational coordinates, and vice-versa, every pseudo-marginal vector with rational coordinates has at least one pre-image in some finite graph cover. (Actually, the papers [4], [5] focused on a special case of graphical models, namely graphical models that represent binary low-density parity-check codes, and consequently dealt with pseudo-codewords. However, the results therein are easily generalized to the more general setup considered here.) The present paper discusses the following extension of this result. Namely, letting β be a pseudo-marginal vector with only rational coordinates and letting $\bar{C}_M(\beta)$ be the average number of pre-images of β among all the M -covers, we show that $\bar{C}_M(\beta)$ grows, when M goes to infinity, like

$$\bar{C}_M(\beta) = \exp(M \cdot H_B(\beta) + o(M)).$$

This characterization of the Bethe entropy function has clearly a “combinatorial flavor,” which is in contrast to the “analytical flavor” of the original definition of the Bethe entropy function in [3] (see Definition 14 in the present paper).

B. A Combinatorial Characterization of the Bethe Partition Function in Terms of Finite Graph Covers

This paper offers also a combinatorial characterization of the Bethe partition function Z_B in terms of finite graph covers of the factor graph under consideration. This is again in contrast to the original, analytical, definition of Z_B that defines Z_B via the minimum of F_B . Compare this with the Gibbs partition function Z_G : its definition is combinatorial in the sense that Z_G is defined as a sum of certain terms. (Of course, the Gibbs partition function can also be characterized analytically via the minimum of the Gibbs free energy function F_G .)

More precisely, recall that the Gibbs partition function (or total sum) of a factor graph N is

$$Z_G(N) \triangleq \sum_{a \in A} g(a)^{1/T},$$

where the sum is over all configurations of N and where g is the global function of N . We show that the Bethe partition function can be written as follows

$$Z_B(N) = \limsup_{M \rightarrow \infty} Z_{B,M}(N), \quad (1)$$

where

$$Z_{B,M}(N) \triangleq \sqrt[M]{\left\langle Z_G(\tilde{N}) \right\rangle_{\tilde{N} \in \tilde{N}_M}}. \quad (2)$$

Here the expression under the root sign represents the average of $Z_G(\tilde{N})$ over all M -covers \tilde{N} of N . Clearly, the expression for $Z_B(N)$ given by (1)–(2) has a “combinatorial flavor.”

Interestingly, the expression in (2) is based on only two rather simple concepts (besides the standard mathematical concepts of taking limits, taking roots, and computing averages): we only need to define the concept of an M -cover of a factor graph and the concept of the Gibbs partition function of a factor graph. In our opinion, these concepts are quite a bit simpler than the ones needed for defining F_B and then $Z_B(N)$ in terms of the minimum of F_B . (A technical note on the side:

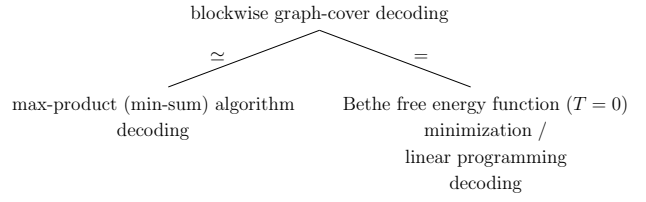


Fig. 1. Blockwise graph-cover decoding forms a bridge between max-product (min-sum) algorithm decoding and linear programming decoding, the latter being equivalent to Bethe free energy minimization at temperature $T = 0$. (See Section I-C for more details.)

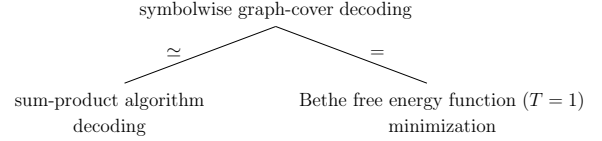


Fig. 2. Symbolwise graph-cover decoding forms a bridge between sum-product algorithm decoding and Bethe free energy minimization at temperature $T = 1$. (See Section I-C for more details.)

in order for the minimum of F_B to make sense, we need the assumption that was stated in Footnote 1, namely that all local functions of N are assumed to be non-negative. As we will see, this assumption is also crucial for showing the equivalence of the left- and right-hand sides of (1).)

Note that (2) contains a finite sum. For small factor graphs N and small M this fact can be exploited, for example, to perform some brute-force computations and come up with conjectures of the relationship of $Z_{B,M}(N)$ with respect to (w.r.t.) $Z_G(N)$. Afterwards, one can try to analytically prove these conjectures about $Z_{B,M}(N)$ for any finite M , and thereby prove a similar conjecture for $Z_B(N)$. This line of reasoning is especially interesting if the proofs can be extended to hold for any factor graph within some class of factor graphs.

C. Graph-Cover Decoding

One of the main motivations of the papers [4], [5] to study finite graph covers of a factor graph N was the fact that finite graph covers of N look locally the same as N . Consequently, any locally operating algorithm, like the max-product algorithm or the sum-product algorithm, “cannot distinguish” if they are operating on N or, implicitly, on any of its finite graph covers. Clearly, for factor graphs *with* cycles, this “non-distinguishability” observation implies fundamental limitations on the conclusions that can be reached by locally operating algorithms because finite graph covers of such factor graphs are “non-trivial” in the sense that they contain valid configurations that “cannot be explained” by valid configurations in the base factor graph. This is in sharp contrast to factor graphs *without* cycles: all M -covers of such factor graphs are “trivial” in the sense that they consist of M independent copies of the base factor graph, and so the set of valid configuration of any M -cover equals the M -fold Cartesian product of the set of valid configurations of the base factor graph with itself.

In fact, in the context of message-passing iterative decoding of graph-based codes *with* cycles, we argued in [5], [6] that these fundamental limitations of message-passing iterative

decoding, in particular of max-product (min-sum) algorithm decoding, imply that these decoders behave less like blockwise maximum a-posteriori decoding (which is equivalent to minimizing the Gibbs free energy function at temperature $T = 0$), but much more like linear programming decoding [7], [8] (which is equivalent to minimizing the Bethe free energy function at temperature $T = 0$). This was done with the help of a theoretical tool called blockwise graph-cover decoding, a tool that in [5] was simply called graph-cover decoding. Namely (see Fig. 1),

- on the one hand we showed the equivalence of blockwise graph-cover decoding and linear programming decoding,
- on the other hand we argued that blockwise graph-cover decoding is a good “model” for the behavior of the max-product (min-sum) algorithm decoding.

This latter connection, namely between blockwise graph-cover decoding and max-product (min-sum) algorithm decoding, is in general only an approximate one. However, in all cases where analytical tools are known that exactly characterize the behavior of max-product algorithm decoding, the connection between blockwise graph-cover decoding and the max-product (min-sum) algorithm decoding is exact.

In this paper, we define symbolwise graph-cover decoding, which tries to capture the essential limitations of sum-product algorithm decoding. It will allow us to argue that for graph-based codes with cycles, sum-product algorithm decoding behaves less like symbolwise maximum a-posteriori decoding (which is equivalent to minimizing the Gibbs free energy function at temperature $T = 1$), but much more like an algorithm that minimizes (at least locally) the Bethe free energy function at temperature $T = 1$. This will be done (see Fig. 2),

- on the one hand, by showing that symbolwise graph-cover decoding is equivalent to (globally) minimizing the Bethe free energy function at temperature $T = 1$,
- and by arguing that symbolwise graph-cover decoding is a good “model” for the behavior of sum-product algorithm decoding.

Similar to the above discussion about blockwise graph-cover decoding, this latter connection, namely between symbolwise graph-cover decoding and sum-product algorithm decoding, is in general an approximate one. However, in many cases where analytical tools are known that exactly characterize the behavior of sum-product algorithm decoding, the connection between symbolwise graph-cover decoding and sum-product algorithm decoding is exact.

In any case, using the combinatorial characterization of the Bethe entropy mentioned earlier in this introduction, one can state that a fixed point of the sum-product algorithm corresponds to a certain pseudo-marginal vector of the factor graph under consideration: it is, after taking a biasing channel-output-dependent term properly into account, the pseudo-marginal vector that has (locally) an extremal number of pre-images in all M -covers, when M goes to infinity.

D. The Shape of the Bethe Entropy Function

The paper concludes with a section on the concavity, or the lack thereof, of the Bethe entropy function. Recall that the

Gibbs entropy function is a concave function, and therefore the Gibbs free energy function is a convex function. However, the Bethe entropy function of factor graphs with cycles does in general not exhibit this property. In fact, in this paper we show that the factor graph associated with any code picked from Gallager’s ensemble of (d_L, d_R) -regular low-density parity-check codes has a Bethe entropy function that is convex in certain regions of its domain if the ensemble is such that the minimum Hamming distance of its codes grows (with high probability) linearly with the block length. This means that there is a trade-off between two desirable objectives: on the one hand to pick a code from a code ensemble with linearly growing minimum Hamming distance, on the other hand to pick a code whose factor graph has a concave Bethe entropy function, *i.e.*, a convex Bethe free energy function.

E. Related Work

Let us briefly discuss some work that is related to the content of this paper.

- Of course, what is called the Bethe approximation in the context of factor graphs has a long history in physics and goes back to ideas that were presented in a 1935 paper by Bethe [9] (see also the 1936 paper by Peierls [10]). Bethe’s approximation therein was mostly an assumption about the conditional independence between different sites in a crystalline alloy. Kurata, Kikuchi, and Watari [11] later on pointed out that this approximation is exact on what they called a “Bethe lattice.” (For further information on these and related topics in physics, we recommend, *e.g.*, [12]–[14].)

Let us comment on the Bethe lattice of a lattice. Consider a lattice L and a factor graph N . In factor-graph language, if L corresponds to N , then the Bethe lattice \hat{L} of L corresponds to the universal cover \hat{N} of N , *i.e.*, the limit of a computation tree of N with arbitrary root in the limit of infinitely many iterations. With the help of \hat{N} it is possible to give a combinatorial characterization of the Bethe partition function of N as some suitably normalized sum of the global function of \hat{N} over all its configurations. However, to make this rigorous, one has to formulate \hat{N} as a suitable limit of computation trees. This is not too difficult for very regular factor graphs or for factor graphs with suitable correlation decay properties. However, for general factor graphs the limit of the above-mentioned normalized sum is rather non-trivial. This difficulty is not quite surprising given, among other reasons, the fact that the SPA may asymptotically exhibit many different types of behaviors (fixed point, periodic, or even “chaotic”), the fact that copies of factor-graph nodes have different multiplicities in finite-size computation trees (see, *e.g.*, [15]), or the fact that the fraction of leaf nodes among all nodes in a computation is non-vanishing in the limit of infinitely many iterations. Clearly, the expression in (1) also contains a limit, however, in our opinion that limit is significantly simpler. Moreover, many effects that are responsible for the similarities and differences between the Bethe and the Gibbs partition function are already visible in finite graph covers with small cover degree.

- Some computations that we will perform in the present paper are very similar to the computations that are necessary to derive the asymptotic growth rate of the average Hamming weight enumerator of proto-graph-based ensembles of (generalized) low-density parity-check (LDPC) codes [16]–[22] (see also the earlier work on uniform interleavers [23], [24]). However, besides some brief mention of the fundamental polytope in [16], these papers do not seem to elaborate on the connection of their results to the Bethe entropy function. (An exception is the very recent paper [22].)
- As already stated in the above introduction, the papers [4], [5] investigated some fundamental limitations that locally operating algorithms have compared to globally operating algorithms. It is worthwhile to point out that Angluin [25], in a paper that was published in 1980, used a very similar global-vs.-local argument to characterize networks of processors. Although on a philosophical level the starting point of her argument is very akin to the one in [4], [5], her conclusions are quite different in nature (which is not so surprising given the differences between her setup and our setup). Much closer to the approach in [4], [5] is the relatively recent paper by Ruozzi *et al.* [26] which showed that results on the limitations of locally operating algorithms on Gaussian graphical models (in particular results based on the concept of walk-summability [27]) can be re-derived by studying graph covers.
- There are other papers where the Bethe approximation plays a central role in characterizing the suboptimal behavior of locally operating algorithms, in particular let us mention [28]–[33].
- Although some concepts in the present paper are evocative of concepts of the replica method (see, *e.g.*, [34] and [35, Chapter 8], [36, Appendix I]), there are also stark differences, as will be discussed in Section VII-E. However, inspired by an earlier version of the present paper, Mori [37] recently showed an alternative (and simpler) approach to some computations that are done in the context of the replica method. See also the follow-up papers [38], [39].
- Let θ be a non-negative square matrix and let $\text{perm}(\theta)$ be the permanent of this matrix [40]. In the paper [41], many concepts of the present paper are specialized to a certain graphical model $N(\theta)$ for which $Z_G(N(\theta)) = \text{perm}(\theta)$ holds. The reformulation of the Bethe partition function for such graphical models was subsequently used by Smarandache [42] to give a proof for a conjecture about pseudo-codewords of LDPC codes. After the initial submission of the present paper and of [41], we became aware of a paper by Greenhill, Janson, and Ruciński [43] that, in the language of the present paper, introduces a graphical model $N'(\theta)$ for which $Z_G(N'(\theta)) = \text{perm}(\theta)$ holds and for which they compute high-order approximations of $Z_{B,M}(N'(\theta))$. However, note that $N'(\theta)$ is in general different from $N(\theta)$. A detailed discussion of connections between $N(\theta)$ and $N'(\theta)$ is given in [41, Section VII.E].

- Based on the reformulation of the Bethe partition function in an earlier version of the present paper, Watanabe [44] stated a conjecture about the relationship of the number of independent sets of a graph and its Bethe approximation (along with other similar conjectures), and Ruozzi [45] proved that the Gibbs partition function of a graphical model with log-supermodular function nodes is always lower bounded by its Bethe partition function, thereby proving a conjecture by Sudderth, Wainwright, and Will-sky [46].
- Graph covers were used in the recent paper [47] to explain why the Bethe partition function is very close to the Gibbs partition function for certain graphical models that appear in the context of constrained coding. They were also used in [48] to prove properties of the Bethe approximation of the so-called pattern maximum likelihood distribution.

F. Overview of the Paper

This paper is structured as follows. We conclude this first section with a subsection on notations and definitions. Then, in Section II we review the basics of normal factor graphs, *i.e.*, the type of factor graphs that we will use in this paper, and in Section III we discuss the Gibbs free energy function and related functions. The Gibbs free energy function is again the topic of Section IV, where we present a simple setup where this function arises naturally. Afterwards, in Section V we move on to introduce the Bethe approximation and the functions that come with it.

After reviewing the main facts about graph covers in Section VI, we come to the main part of this paper, namely Section VII where we present the promised combinatorial characterization of the Bethe entropy function and the Bethe partition function.

In contrast to the previous sections that considered a general factor graph setup, the next three sections focus on factor graphs that appear in coding theory. Namely, Section VIII reviews some relevant concepts, Section IX discusses block-wise and symbolwise graph-cover decoding, and Section X investigates the influence of the minimum Hamming distance of a code upon the Bethe entropy function of its factor graph.

Finally, the paper is concluded in Section XI. The longer proofs of the lemmas and theorems in the main text are collected in the appendices.

G. Basic Notations and Definitions

This subsection discusses the most important notations that will be used in this paper. More notational definitions will be given in later sections.

We let \mathbb{Z} , $\mathbb{Z}_{\geq 0}$, $\mathbb{Z}_{> 0}$, \mathbb{R} , $\mathbb{R}_{\geq 0}$, $\mathbb{R}_{> 0}$, and \mathbb{F}_2 be, respectively, the ring of integers, the set of non-negative integers, the set of positive integers, the field of real numbers, the set of non-negative real numbers, the set of positive real numbers, and the Galois field of size two. Scalars are denoted by non-boldface characters, whereas vectors and matrices by boldface characters. All logarithms in this paper will be natural logarithms, and so entropies will be measured in nats. (The only

exception are figures where entropies are shown in bits.) As usually done in information theory, we define $\log(0) \triangleq -\infty$ and $0 \cdot \log(0) \triangleq 0$.

Sets are denoted by calligraphic letters, and the size of a finite set \mathcal{S} is written like $|\mathcal{S}|$. The convex hull and the conic hull [49] of some set $\mathcal{R} \in \mathbb{R}^n$ are, respectively, denoted by $\text{conv}(\mathcal{R})$ and $\text{conic}(\mathcal{R})$.

We use square brackets in two different ways. Namely, for any $L \in \mathbb{Z}_{>0}$ we define $[L] \triangleq \{1, \dots, L\}$, and for any statement S we follow Iverson's convention by defining $[S] \triangleq 1$ if S is true and $[S] \triangleq 0$ otherwise.

Finally, for a finite set \mathcal{S} , we define $\Pi_{\mathcal{S}}$ to be the set of vectors representing probability mass functions over \mathcal{S} , i.e.,

$$\Pi_{\mathcal{S}} \triangleq \left\{ \mathbf{p} = (p_s)_{s \in \mathcal{S}} \mid p_s \geq 0 \text{ for all } s \in \mathcal{S}, \sum_{s \in \mathcal{S}} p_s = 1 \right\}.$$

II. NORMAL FACTOR GRAPHS

Factor graphs are a convenient way to represent multivariate functions [50]. In this paper we use a variant called normal factor graphs (NFGs) [51] (also called Forney-style factor graphs [52]), where variables are associated with edges.

The key aspects of an NFG are best explained with the help of an example.

Example 1 Consider the multivariate function

$$\begin{aligned} g(a_{e_1}, \dots, a_{e_8}) &\triangleq g_{f_1}(a_{e_1}, a_{e_2}, a_{e_5}) \cdot g_{f_2}(a_{e_2}, a_{e_3}, a_{e_6}) \\ &\quad \cdot g_{f_3}(a_{e_3}, a_{e_4}, a_{e_7}) \cdot g_{f_4}(a_{e_5}, a_{e_6}, a_{e_8}) \\ &\quad \cdot g_{f_5}(a_{e_7}, a_{e_8}), \end{aligned}$$

where the so-called global function g is the product of the so-called local functions g_{f_1} , g_{f_2} , g_{f_3} , g_{f_4} , and g_{f_5} . The decomposition of this global function as a product of local functions can be depicted with the help of an NFG \mathbf{N} as shown in Fig. 3. In particular, the NFG \mathbf{N} consists of

- the function nodes f_1 , f_2 , f_3 , f_4 , and f_5 ;
- the half-edges e_1 and e_4 (sometimes also called “external edges”);
- the full-edges e_2 , e_3 , e_5 , e_6 , e_7 , and e_8 (sometimes also called “internal edges”).

In general,

- a function node f represents the local function g_f ;
- with an edge e we associate the variable A_e (note that a realization of the variable A_e is denoted by a_e);
- an edge e is incident on a function node f if and only if a_e appears as an argument of the local function g_f .

Note that the NFG \mathbf{N} contains three cycles, one involving the edges e_2 , e_5 , e_6 , one involving the edges e_3 , e_6 , e_7 , e_8 , and one involving the edges e_2 , e_3 , e_5 , e_7 , e_8 . As is well known from the literature on graphical models, and as we can also see from other parts of this paper, the existence/absence of cycles in an NFG has significant implications for its properties, in particular with respect to the behavior of locally operating algorithms like the max-product algorithm and the sum-product algorithm. \square

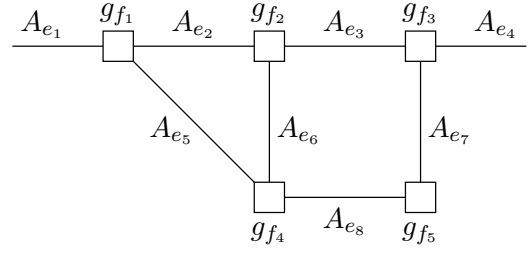


Fig. 3. NFG \mathbf{N} used in Example 1.

We now present the general definition of an NFG that we will use in this paper.

Definition 2 An NFG $\mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ consists of the following objects.

- A graph $(\mathcal{F}, \mathcal{E})$ with vertex set \mathcal{F} (also known as the function node set) and with edge set $\mathcal{E} = \mathcal{E}_{\text{half}} \cup \mathcal{E}_{\text{full}}$, where $\mathcal{E}_{\text{half}}$ and $\mathcal{E}_{\text{full}}$ represent, respectively, the set of half-edges and the set of full-edges.
- A collection of alphabets $\mathcal{A} \triangleq \{A_e\}_{e \in \mathcal{E}}$, where the alphabet A_e is associated with the edge $e \in \mathcal{E}$. In the following, with a slight abuse of notation, \mathcal{A} will also stand for the set $\mathcal{A} = \prod_{e \in \mathcal{E}} A_e$, i.e., the Cartesian product of the alphabets $\{A_e\}_{e \in \mathcal{E}}$. Moreover, we also define the sets $\mathcal{A}_{\text{half}} \triangleq \prod_{e \in \mathcal{E}_{\text{half}}} A_e$ and $\mathcal{A}_{\text{full}} \triangleq \prod_{e \in \mathcal{E}_{\text{full}}} A_e$. Clearly, $\mathcal{A} = \mathcal{A}_{\text{half}} \times \mathcal{A}_{\text{full}}$.
- A collection of functions $\mathcal{G} \triangleq \{g_f\}_{f \in \mathcal{F}}$ (called local functions), where the function g_f is associated with the function node $f \in \mathcal{F}$ and further specified below. \square

Definition 3 Given an NFG $\mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$, we make the following definitions.

- For every $f \in \mathcal{F}$, we define \mathcal{E}_f to be the set of edges incident on f .
- A vector \mathbf{a} in the set \mathcal{A} , i.e.,

$$\mathbf{a} \triangleq (a_e)_{e \in \mathcal{E}} \in \mathcal{A},$$

will be called a configuration of the NFG. For a given vector \mathbf{a} , we also define for every $f \in \mathcal{F}$ the sub-vector

$$\mathbf{a}_f \triangleq (a_{f,e})_{e \in \mathcal{E}_f} \triangleq (a_e)_{e \in \mathcal{E}_f}.$$

Note that we will also use the notation $\mathbf{a}_f = (a_{f,e})_{e \in \mathcal{E}_f}$ when there is not necessarily an underlying configuration \mathbf{a} of the whole NFG.

- For every $f \in \mathcal{F}$, the local function g_f is an arbitrary mapping

$$g_f : \prod_{e \in \mathcal{E}_f} A_e \rightarrow \mathbb{R}, \quad \mathbf{a}_f \mapsto g_f(\mathbf{a}_f).$$

- For every $f \in \mathcal{F}$, we define the function node alphabet \mathcal{A}_f to be the set

$$\mathcal{A}_f \triangleq \left\{ \mathbf{a}_f \in \prod_{e \in \mathcal{E}_f} A_e \mid g_f(\mathbf{a}_f) \neq 0 \right\}.$$

III. THE GIBBS FREE ENERGY FUNCTION AND THE GIBBS PARTITION FUNCTION

This section reviews the concept of the Gibbs free energy function of an NFG, along with some related functions. The temperature T appears in them as a parameter.

Assumption 6 *Throughout this paper, the temperature T will be some fixed non-negative real number, i.e., $T \in \mathbb{R}_{\geq 0}$. (If a definition requires $T \in \mathbb{R}_{>0}$, then this will be pointed out.)* \square

The Gibbs free energy function is defined such that its minimal value, along with the location of its minimal value, encode important information about the global function that is represented by the NFG. In particular, for $T \in \mathbb{R}_{>0}$ the minimal value equals the temperature T times the negative logarithm of the partition function, where the partition function is the sum of the $(1/T)$ -th power of the global function over all configurations.

Definition 7 *Consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$. For any temperature $T \in \mathbb{R}_{\geq 0}$, the Gibbs free energy function associated with N is defined to be (see, e.g., [3])*

$$F_G : \Pi_C \rightarrow \mathbb{R}, \quad \mathbf{p} \mapsto U_G(\mathbf{p}) - T \cdot H_G(\mathbf{p}),$$

where

$$U_G : \Pi_C \rightarrow \mathbb{R}, \quad \mathbf{p} \mapsto - \sum_{\mathbf{c}} p_{\mathbf{c}} \cdot \log(g(\mathbf{c})),$$

$$H_G : \Pi_C \rightarrow \mathbb{R}, \quad \mathbf{p} \mapsto - \sum_{\mathbf{c}} p_{\mathbf{c}} \cdot \log(p_{\mathbf{c}}).$$

Here, U_G is called the Gibbs average energy function and H_G is called the Gibbs entropy function.

Moreover, for $T \in \mathbb{R}_{>0}$, the (Gibbs) partition function associated with N is defined to be (see, e.g., [3])

$$Z_G \triangleq \sum_{\mathbf{a} \in \mathcal{A}} g(\mathbf{a})^{1/T} = \sum_{\mathbf{c} \in \mathcal{C}} g(\mathbf{c})^{1/T}. \quad (4)$$

\square

Note that “function” in “partition function” refers to the fact that the expression in (4) typically is a function of some parameters like the temperature T . (A better word for “partition function” would possibly be “partition sum” or “state sum,” which would more closely follow the German “Zustandssumme” whose first letter is used to denote the partition function.)

Lemma 8 *Consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$. For $T \in \mathbb{R}_{\geq 0}$, the function $F_G(\mathbf{p})$ is convex in \mathbf{p} . Moreover, for $T \in \mathbb{R}_{>0}$, the function $F_G(\mathbf{p})$ is minimized by $\mathbf{p} = \mathbf{p}^*$, where*

$$\mathbf{p}_{\mathbf{c}}^* \triangleq \frac{g(\mathbf{c})^{1/T}}{Z_G}, \quad \mathbf{c} \in \mathcal{C}. \quad (5)$$

At its minimum, F_G takes on the value

$$F_G(\mathbf{p}^*) = -T \cdot \log(Z_G), \quad (6)$$

which is also known as the Helmholtz free energy.

Proof: For $T = 0$ we have $F_G(\mathbf{p}) = U_G(\mathbf{p})$, $\mathbf{p} \in \Pi_C$. The convexity of F_G then follows from the convexity of U_G ,

which follows from the fact that U_G is a linear function of its argument.

For $T \in \mathbb{R}_{>0}$, the Gibbs free energy function F_G can be expressed in terms of a relative entropy functional, namely

$$F_G(\mathbf{p}) = T \cdot D \left((p_{\mathbf{c}})_{\mathbf{c} \in \mathcal{C}} \parallel \left(\frac{g(\mathbf{c})^{1/T}}{Z_G} \right)_{\mathbf{c} \in \mathcal{C}} \right) - T \cdot \log(Z_G).$$

The statements in the lemma then follow easily from standard properties of the relative entropy functional (see, e.g., [53]). \blacksquare

Note that, with appropriate care, results involving the Gibbs free energy function and related functions at temperature $T = 0$ can be recovered from studying the case $T \in \mathbb{R}_{>0}$ and taking the limit $T \downarrow 0$. However, as mentioned in Assumption 6, in this paper the temperature T is a fixed parameter and we will not consider such limits.

Let us briefly discuss a variant of the above Gibbs free energy function. Namely, for some $\mathbf{c}_{\text{half}} \in \mathcal{C}_{\text{half}}$, we will say that $\mathbf{p} \in \Pi_C$ is compatible with \mathbf{c}_{half} if

$$\mathbf{p}_{\mathbf{c}'} \begin{cases} \geq 0 & (\text{for all } \mathbf{c}' \in \mathcal{C} \text{ with } \mathbf{c}'_{\text{half}} = \mathbf{c}_{\text{half}}), \\ = 0 & (\text{for all } \mathbf{c}' \in \mathcal{C} \text{ with } \mathbf{c}'_{\text{half}} \neq \mathbf{c}_{\text{half}}) \end{cases}.$$

With this definition, as an alternative to the minimization problem in Lemma 8, we can consider a minimization problem where we minimize over all \mathbf{p} that are compatible with some given \mathbf{c}_{half} . Technically, we can accomplish this by defining a modified Gibbs free energy function F'_G that equals the Gibbs free energy function F_G for \mathbf{p} 's which are compatible with this \mathbf{c}_{half} , and that is infinite otherwise.

Definition 9 *Consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$. For any temperature $T \in \mathbb{R}_{\geq 0}$, the modified Gibbs free energy function associated with N is defined to be*

$$F'_G : \mathcal{C}_{\text{half}} \times \Pi_C \rightarrow \mathbb{R} \cup \{+\infty\}$$

$$(\mathbf{c}_{\text{half}}, \mathbf{p}) \mapsto \begin{cases} F_G(\mathbf{p}) & (\mathbf{p} \text{ is compatible with } \mathbf{c}_{\text{half}}) \\ +\infty & (\text{otherwise}) \end{cases}$$

Moreover, for $T \in \mathbb{R}_{>0}$, the modified (Gibbs) partition function associated with N is defined to be

$$Z'_G : \mathcal{C}_{\text{half}} \rightarrow \mathbb{R}, \quad \mathbf{c}_{\text{half}} \mapsto \sum_{\mathbf{c}' \in \mathcal{C}: \mathbf{c}'_{\text{half}} = \mathbf{c}_{\text{half}}} g(\mathbf{c}')^{1/T}.$$

Lemma 10 *Consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ and fix some $\mathbf{c}_{\text{half}} \in \mathcal{C}_{\text{half}}$. For $T \in \mathbb{R}_{\geq 0}$, the function $F'_G(\mathbf{c}_{\text{half}}, \mathbf{p})$ is convex in \mathbf{p} . Moreover, for $T \in \mathbb{R}_{>0}$, the function $F'_G(\mathbf{c}_{\text{half}}, \mathbf{p})$ is minimized by $\mathbf{p} = \mathbf{p}^*$, where*

$$\mathbf{p}_{\mathbf{c}'}^* \triangleq \begin{cases} \frac{g(\mathbf{c}')^{1/T}}{Z'_G(\mathbf{c}_{\text{half}})}, & (\mathbf{c}' \in \mathcal{C} \text{ with } \mathbf{c}'_{\text{half}} = \mathbf{c}_{\text{half}}), \\ 0 & (\text{otherwise}) \end{cases}.$$

At its minimum, $F'_G(\mathbf{c}_{\text{half}}, \cdot)$ takes on the value

$$F'_G(\mathbf{c}_{\text{half}}, \mathbf{p}^*) = -T \cdot \log(Z'_G(\mathbf{c}_{\text{half}})).$$

Proof: Similar to the proof of Lemma 8. \blacksquare

Clearly, if the NFG N does not contain any half-edges, then the modified Gibbs partition function is essentially a

scalar. Note that this modified Gibbs partition function (at temperature $T = 1$) equals what Forney [54] calls the partition function of an NFG and what Al-Bashabsheh and Mao [55] call the exterior function of an NFG.

IV. WHY THE GIBBS FREE ENERGY FUNCTION ARISES RATHER NATURALLY

Of course, besides the function Gibbs free energy function F_G , there are many ways to formulate a function $F : \Pi_C \rightarrow \mathbb{R}$ such that

- the minimum of $F(\mathbf{p})$ is achieved at $\mathbf{p} = \mathbf{p}^*$, where \mathbf{p}^* equals the expression in (5),
- and such that the minimum value $F(\mathbf{p}^*)$ equals the expression in (6).

The goal of this section is to discuss a setup where F_G arises rather naturally as a function that has these properties. This section also reviews some important concepts from the method of types (see, e.g., [53], [56]), many of which inspired the concepts that were briefly mentioned in Section I and that will be introduced more thoroughly in Sections VI and VII. Note that we do not make any novelty claims for the observations discussed in the present section.

Throughout this section we consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$. For simplicity we consider only the case where $\mathcal{A}_e = \{0, 1\} \subset \mathbb{R}$ for all $e \in \mathcal{E}$. Moreover, throughout this discussion, the temperature will be fixed to $T = 1$. (Both assumptions are not critical, and with a suitable formalism, the results can easily be generalized.)

We start by defining a probability mass function $P_C(\mathbf{c})$ on \mathcal{C} that is induced by the global function on N , namely

$$P_C(\mathbf{c}) \triangleq \frac{g(\mathbf{c})}{Z_G}, \quad \mathbf{c} \in \mathcal{C},$$

where Z_G is defined in (4). Now, assume that for every $e \in \mathcal{E}$ and $a_e \in \mathcal{A}_e$ we want to compute the marginal

$$P_{A_e}(a_e) \triangleq \sum_{\mathbf{c}: c_e = a_e} P_C(\mathbf{c}).$$

(This computational problem comes up, for example, as part of symbolwise maximum a-posteriori decoding, cf. Section IX-C.) To that end, define the vector $\boldsymbol{\eta} \triangleq (\eta_{e,1})_{e \in \mathcal{E}}$ with components $\eta_{e,1} \triangleq P_{A_e}(1)$. Clearly, once we have computed the vector $\boldsymbol{\eta}$, we have all the desired marginals because $P_{A_e}(0) = 1 - P_{A_e}(1) = 1 - \eta_{e,1}$, $e \in \mathcal{E}$. It can easily be verified that $\boldsymbol{\eta}$ satisfies

$$\boldsymbol{\eta} = \sum_{\mathbf{c} \in \mathcal{C}} P_C(\mathbf{c}) \cdot \mathbf{c} = \mathbb{E}[\mathbf{C}]. \quad (7)$$

In practice, the sum in (7) is very often intractable because the set \mathcal{C} is very large. However, (7) shows that $\boldsymbol{\eta}$ is some expectation value and so we can try to approximate it by stochastic averaging. Therefore, let

$$\mathbf{c}_{\text{sample}(1)}, \mathbf{c}_{\text{sample}(2)}, \dots, \mathbf{c}_{\text{sample}(M)}$$

be M i.i.d. sample vectors distributed according to P_C . Then

$$\boldsymbol{\eta} \approx \frac{1}{M} \sum_{m \in [M]} \mathbf{a}_{\text{sample}(m)}. \quad (8)$$

(This expression is akin to the expression in [35, Section 13.2] on “decoding by sampling.”) This approach can work well for certain NFGs. However, in general it is, unfortunately, difficult to efficiently obtain enough i.i.d. samples so that $\boldsymbol{\eta}$ can be estimated with sufficient accuracy. The difficulty of generating i.i.d. samples happens for example in the case of NFGs that represent good codes, see the discussion in [35, Section 13.2]. Therefore, the expression in (8) does not offer a shortcut for the main computational step in symbolwise maximum a-posteriori decoding. (Of course, this observation is not really surprising given the well-known computational complexity of that decoder.)

Nevertheless, conceptually the expression in (8) is very useful as it suggests the following considerations that will lead to a function that fulfills the promises that were stated at the beginning of this section. Namely, let

$$\tilde{\mathbf{C}}_1, \tilde{\mathbf{C}}_2, \dots, \tilde{\mathbf{C}}_M$$

be M i.i.d. random vectors with distribution P_C . Then

$$\begin{aligned} \boldsymbol{\eta} = \mathbb{E}[\mathbf{C}] &\stackrel{(a)}{=} \frac{1}{M} \sum_{m \in [M]} \mathbb{E}[\tilde{\mathbf{C}}_m] = \mathbb{E} \left[\frac{1}{M} \sum_{m \in [M]} \tilde{\mathbf{C}}_m \right] \\ &= \sum_{\tilde{\mathbf{c}}_1 \in \mathcal{C}} \cdots \sum_{\tilde{\mathbf{c}}_M \in \mathcal{C}} \left(\prod_{m \in [M]} P_C(\tilde{\mathbf{c}}_m) \right) \frac{1}{M} \sum_{m \in [M]} \tilde{\mathbf{c}}_m, \end{aligned} \quad (9)$$

where step (a) follows trivially from the definition of $\tilde{\mathbf{C}}_m$, $m \in [M]$, and was inspired by (8). As simple as it is, this step is actually the only “non-trivial” step in the whole discussion here. The rest will simply be a “mechanical” application of the method of types (see, e.g., [53], [56]) towards simplifying the expression in (9).

Therefore, let us recall the relevant definitions from the method of types.

Definition 11 Consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ and fix some integer $M \in \mathbb{Z}_{>0}$.

- **(Mapping)** Define the mapping

$$\varphi_M : \mathcal{C}^M \rightarrow \Pi_C, \quad \tilde{\mathbf{c}} \triangleq (\tilde{\mathbf{c}}_m)_{m \in [M]} \mapsto \mathbf{q}^{(\tilde{\mathbf{c}})},$$

where

$$\begin{aligned} q_c^{(\tilde{\mathbf{c}})} &\triangleq \frac{1}{M} \cdot (\text{number of appearances of } \mathbf{c} \text{ in } \tilde{\mathbf{c}}) \\ &= \frac{1}{M} \sum_{m \in [M]} [\tilde{\mathbf{c}}_m = \mathbf{c}], \quad \mathbf{c} \in \mathcal{C}. \end{aligned}$$

(In the above expression we have used Iverson’s convention that was defined in Section I-G.)

- **(Type)** Let $\tilde{\mathbf{c}}$ be a sequence over \mathcal{C} of length M , i.e., $\tilde{\mathbf{c}} \triangleq (\tilde{\mathbf{c}}_m)_{m \in [M]} \in \mathcal{C}^M$. Then the vector

$$\mathbf{q}^{(\tilde{\mathbf{c}})} \triangleq \varphi_M(\tilde{\mathbf{c}})$$

is called the type of $\tilde{\mathbf{c}}$, or the empirical probability distribution of $\tilde{\mathbf{c}}$.

- **(Set of all possible types)** The set $\mathcal{Q}_M \subseteq \Pi_C$ is defined to be the set of all possible types that are based on sequences over \mathcal{C} of length M , i.e.,

$$\mathcal{Q}_M \triangleq \varphi_M(\mathcal{C}^M).$$

- **(Type class)** For any $\mathbf{q} \in \mathcal{Q}_M$, the type class of \mathbf{q} is defined to be the set of all vectors in \mathcal{C}^M with type \mathbf{q} . Equivalently, the type class of \mathbf{q} is defined to be the pre-image of \mathbf{q} under the mapping φ_M , i.e.,

$$\mathcal{T}_M(\mathbf{q}) \triangleq \varphi_M^{-1}(\mathbf{q}) = \left\{ \tilde{\mathbf{c}} \in \mathcal{C}^M \mid \mathbf{q}^{(\tilde{\mathbf{c}})} = \mathbf{q} \right\}.$$

- **(Mean vector)** Assume $\mathcal{A}_e = \{0, 1\} \subset \mathbb{R}$ for all $e \in \mathcal{E}$. For any type $\mathbf{q} \in \mathcal{Q}_M$, the mean vector associated with \mathbf{q} is defined to be

$$\text{mean}(\mathbf{q}) \triangleq \sum_{\mathbf{c} \in \mathcal{C}} q_{\mathbf{c}} \cdot \mathbf{c}.$$

□

The following lemma contains some well-known properties of the objects that were introduced in the above definition.

Lemma 12 Consider an NFG $\mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ and fix some integer $M \in \mathbb{Z}_{>0}$.

- The size of the set \mathcal{Q}_M is upper bounded as follows

$$|\mathcal{Q}_M| \leq (M+1)^{|\mathcal{C}|}.$$

Because $|\mathcal{C}|$ is a fixed number for a given NFG \mathbf{N} , this upper bound is a polynomial in M .

- Let $\tilde{\mathbf{c}} \triangleq (\tilde{c}_m)_{m \in [M]} \in \mathcal{C}^M$, i.e., $\tilde{\mathbf{c}}$ is a sequence over \mathcal{C} of length M . Then

$$\prod_{m \in [M]} P_{\mathcal{C}}(\tilde{c}_m) = \frac{1}{Z_G^M} \exp \left(-M \cdot U_G \left(\mathbf{q}^{(\tilde{\mathbf{c}})} \right) \right),$$

where U_G and Z_G are, respectively, the Gibbs average energy function and the Gibbs partition function associated with \mathbf{N} , see Definition 7.

- For any type $\mathbf{q} \in \mathcal{Q}_M$, let $C_M(\mathbf{q})$ be the size of the type class $\mathcal{T}_M(\mathbf{q})$ of \mathbf{q} . Then

$$C_M(\mathbf{q}) = \exp \left(M \cdot H_G(\mathbf{q}) + o(M) \right),$$

where H_G is the Gibbs entropy function associated with \mathbf{N} , see Definition 7. (Because $C_M(\mathbf{q})$ is counting certain objects, this characterization of the Gibbs entropy function has clearly a “combinatorial flavor” which is in contrast to the “analytical flavor” of the Gibbs entropy function in Definition 7.)

- Let $\tilde{\mathbf{c}} \triangleq (\tilde{c}_m)_{m \in [M]} \in \mathcal{C}^M$, i.e., $\tilde{\mathbf{c}}$ is a sequence over \mathcal{C} of length M . Then

$$\frac{1}{M} \sum_{m \in [M]} \tilde{c}_m = \text{mean} \left(\mathbf{q}^{(\tilde{\mathbf{c}})} \right).$$

Proof: See, e.g., [53], [56]. ■

With this, the M -fold summation $\sum_{\tilde{\mathbf{c}}_1} \cdots \sum_{\tilde{\mathbf{c}}_M}$ in (9) can be replaced by the double summation $\sum_{\mathbf{q} \in \mathcal{Q}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{T}_M(\mathbf{q})}$, and

we obtain

$$\begin{aligned} \eta &= \sum_{\mathbf{q} \in \mathcal{Q}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{T}_M(\mathbf{q})} \underbrace{\left(\prod_{m \in [M]} P_{\mathcal{C}}(\tilde{c}_m) \right)}_{\stackrel{(a)}{=} \frac{1}{Z_G^M} \cdot \exp(-M \cdot U_G(\mathbf{q}))} \cdot \underbrace{\frac{1}{M} \sum_{m \in [M]} \tilde{c}_m}_{\stackrel{(b)}{=} \text{mean}(\mathbf{q})} \\ &= \sum_{\mathbf{q} \in \mathcal{Q}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{T}_M(\mathbf{q})} \frac{1}{Z_G^M} \cdot \exp(-M \cdot U_G(\mathbf{q})) \cdot \text{mean}(\mathbf{q}) \\ &\stackrel{(c)}{=} \sum_{\mathbf{q} \in \mathcal{Q}_M} \frac{1}{Z_G^M} \cdot \exp(-M \cdot U_G(\mathbf{q})) \cdot \text{mean}(\mathbf{q}) \cdot \underbrace{\sum_{\tilde{\mathbf{c}} \in \mathcal{T}_M(\mathbf{q})} 1}_{\stackrel{(d)}{=} C_M(\mathbf{q})} \\ &\stackrel{(e)}{=} \sum_{\mathbf{q} \in \mathcal{Q}_M} s_M(\mathbf{q}) \cdot \text{mean}(\mathbf{q}), \end{aligned} \quad (10)$$

where at steps (a) and (b) we have used Lemma 12, where at step (c) we have used the fact that the terms appearing in the summation are independent of $\tilde{\mathbf{c}}$ given their type \mathbf{q} , where at step (d) we have used Lemma 12, and where at step (e) we have used the abbreviation

$$s_M(\mathbf{q}) \triangleq \frac{1}{Z_G^M} \cdot \exp(-M \cdot U_G(\mathbf{q})) \cdot C_M(\mathbf{q}). \quad (11)$$

Similarly, we obtain

$$\begin{aligned} 1 &= \sum_{\tilde{\mathbf{c}}_1} \cdots \sum_{\tilde{\mathbf{c}}_M} \prod_{m \in [M]} P_{\mathcal{C}}(\tilde{c}_m) \\ &= \sum_{\mathbf{q} \in \mathcal{Q}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{T}_M(\mathbf{q})} \underbrace{\prod_{m \in [M]} P_{\mathcal{C}}(\tilde{c}_m)}_{\stackrel{(a)}{=} \frac{1}{Z_G^M} \cdot \exp(-M \cdot U_G(\mathbf{q}))} \\ &= \sum_{\mathbf{q} \in \mathcal{Q}_M} \frac{1}{Z_G^M} \cdot \exp(-M \cdot U_G(\mathbf{q})) \cdot \underbrace{\sum_{\tilde{\mathbf{c}} \in \mathcal{T}_M(\mathbf{q})} 1}_{\stackrel{(d)}{=} C_M(\mathbf{q})} \\ &= \sum_{\mathbf{q} \in \mathcal{Q}_M} s_M(\mathbf{q}), \end{aligned} \quad (12)$$

i.e., s_M is a probability mass function on \mathcal{Q}_M . Moreover, using Lemma 12, it follows from (11) that

$$\begin{aligned} s_M(\mathbf{q}) &= \frac{1}{Z_G^M} \cdot \exp \left(-M \cdot (U_G(\mathbf{q}) - H_G(\mathbf{q})) + o(M) \right) \\ &= \frac{1}{Z_G^M} \cdot \exp \left(-M \cdot F_G(\mathbf{q}) + o(M) \right), \quad \mathbf{q} \in \mathcal{Q}_M. \end{aligned} \quad (13)$$

Because (10) holds for any M , we might as well take the limit $M \rightarrow \infty$. Then, because in the limit $M \rightarrow \infty$ the probability mass function s_M is concentrated more and more around

$$\mathbf{q}^* \triangleq \arg \min_{\mathbf{q} \in \Pi_{\mathcal{C}}} F_G(\mathbf{q}),$$

and because the size of \mathcal{Q}_M grows at most polynomially in M , it follows that in the limit $M \rightarrow \infty$ the sum in (10) can be simplified, and we obtain

$$\eta = \text{mean}(\mathbf{q}^*).$$

Moreover, using (13), the expression in (12) can be rewritten for finite M to read

$$Z_G = \sqrt[M]{\sum_{\mathbf{q} \in \mathcal{Q}_M} \exp(-M \cdot F_G(\mathbf{q}) + o(M))}.$$

Taking the limit $M \rightarrow \infty$, and using the fact that the size of \mathcal{Q}_M grows at most polynomially in M , we can write

$$Z_G = \exp(-F_G(\mathbf{q}^*)) = \exp\left(-\arg \min_{\mathbf{q} \in \Pi_C} F_G(\mathbf{q})\right),$$

i.e.,

$$-\log(Z_G) = \arg \min_{\mathbf{q} \in \Pi_C} F_G(\mathbf{q}).$$

We conclude this section with some remarks.

- Note again that the only “non-trivial” step in the above derivation was step (a) in (9).
- Recall that we assumed that $T = 1$; however, the results in this section can easily be generalized to $T \in \mathbb{R}_{>0}$.
- We could have started this section by defining

$$P_A(\mathbf{a}) \triangleq \frac{g(\mathbf{a})}{Z_G}, \quad \mathbf{a} \in \mathcal{A},$$

and then we could have continued by replacing \mathcal{C} by \mathcal{A} , Π_C by Π_A , etc. (Clearly, $P_A(\mathbf{a}) = P_C(\mathbf{a})$ if $\mathbf{a} \in \mathcal{C}$ and $P_A(\mathbf{a}) = 0$ if $\mathbf{a} \notin \mathcal{C}$.) Both approaches, the approach taken in this section and this alternative approach, have their advantages and disadvantages, but ultimately they yield equivalent results.

V. THE LOCAL MARGINAL POLYTOPE AND THE BETHE APPROXIMATION

In many problems it is desirable to compute the Gibbs partition function of some graphical model. However, the direct evaluation of (4) is usually intractable. Moreover, although the above reformulation of the Gibbs partition function via the minimum value of some function, see (6), is an elegant reformulation of the Gibbs partition function computation problem, this does not yield any computational savings yet.

Nevertheless, it suggests to look for a function that is tractable and whose minimum is close to the minimum of the Gibbs free energy function. An ansatz for such a function is the so-called Bethe free energy function [3]. The Bethe free energy function is interesting because a theorem by Yedidia, Freeman, and Weiss [3] says that fixed points of the sum-product algorithm correspond to stationary points of the Bethe free energy function. (For further motivations for the Bethe approximation we refer to the discussion in [3], [57], [58].)

Before we can state the definition of the Bethe free energy function, we need the concept of the local marginal polytope.

Definition 13 Let $\mathcal{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG and let

$$\beta \triangleq ((\beta_f)_{f \in \mathcal{F}}, (\beta_e)_{e \in \mathcal{E}})$$

be a collection of vectors based on the real vectors

$$\begin{aligned} \beta_f &\triangleq (\beta_{f, \mathbf{a}_f})_{\mathbf{a}_f \in \mathcal{A}_f}, & f \in \mathcal{F} \\ \beta_e &\triangleq (\beta_{e, \mathbf{a}_e})_{\mathbf{a}_e \in \mathcal{A}_e}, & e \in \mathcal{E}. \end{aligned}$$

Then, for $f \in \mathcal{F}$, the f th local marginal polytope (or f th belief polytope) \mathcal{B}_f is defined to be the set

$$\mathcal{B}_f \triangleq \Pi_{\mathcal{A}_f},$$

and for $e \in \mathcal{E}$, the e th local marginal polytope (or e th belief polytope) \mathcal{B}_e is defined to be the set

$$\mathcal{B}_e \triangleq \Pi_{\mathcal{A}_e}.$$

With this, the local marginal polytope (or belief polytope) \mathcal{B} is defined to be the set

$$\mathcal{B} = \left\{ \beta \left| \begin{array}{l} \beta_f \in \mathcal{B}_f \text{ for all } f \in \mathcal{F} \\ \beta_e \in \mathcal{B}_e \text{ for all } e \in \mathcal{E} \\ \sum_{\mathbf{a}'_f \in \mathcal{A}_f: \mathbf{a}'_f, \mathbf{a}_e = \mathbf{a}_e} \beta_{f, \mathbf{a}'_f} = \beta_{e, \mathbf{a}_e} \\ \text{for all } f \in \mathcal{F}, e \in \mathcal{E}_f, \mathbf{a}_e \in \mathcal{A}_e \end{array} \right. \right\},$$

where $\beta \in \mathcal{B}$ is called a pseudo-marginal vector, or more precisely, a locally consistent pseudo-marginal vector. The constraints that were listed last in the definition of \mathcal{B} will be called “edge consistency constraints.” \square

Definition 14 For any temperature $T \in \mathbb{R}_{\geq 0}$, the Bethe free energy function associated with some NFG $\mathcal{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ is defined to be the function (see [3])

$$F_B : \mathcal{B} \rightarrow \mathbb{R}, \quad \beta \mapsto U_B(\beta) - T \cdot H_B(\beta),$$

where

$$U_B : \mathcal{B} \rightarrow \mathbb{R}, \quad \beta \mapsto \sum_f U_{B,f}(\beta_f),$$

$$H_B : \mathcal{B} \rightarrow \mathbb{R}, \quad \beta \mapsto \sum_f H_{B,f}(\beta_f) - \sum_{e \in \mathcal{E}_{\text{full}}} H_{B,e}(\beta_e),$$

with

$$U_{B,f} : \mathcal{B}_f \rightarrow \mathbb{R}, \quad \beta_f \mapsto - \sum_{\mathbf{a}_f} \beta_{f, \mathbf{a}_f} \cdot \log(g_f(\mathbf{a}_f)),$$

$$H_{B,f} : \mathcal{B}_f \rightarrow \mathbb{R}, \quad \beta_f \mapsto - \sum_{\mathbf{a}_f} \beta_{f, \mathbf{a}_f} \cdot \log(\beta_{f, \mathbf{a}_f}),$$

$$H_{B,e} : \mathcal{B}_e \rightarrow \mathbb{R}, \quad \beta_e \mapsto - \sum_{\mathbf{a}_e} \beta_{e, \mathbf{a}_e} \cdot \log(\beta_{e, \mathbf{a}_e}).$$

Here, U_B is the Bethe average energy function and H_B is the Bethe entropy function. \square

Note that in the above definition of $H_B(\beta)$, the term $H_{B,e}(\beta_e)$ appears with coefficient -1 for full-edges $e \in \mathcal{E}_{\text{full}}$, whereas it appears with coefficient 0 for half-edges $e \in \mathcal{E}_{\text{half}}$. Therefore, the latter terms are omitted.²

²These coefficients are consistent with the coefficients in [3]. Namely, because half/full-edges in NFGs correspond to variable nodes of degree one/two in factor graphs [52], and because the Bethe entropy function term corresponding to a degree- d variable node in a factor graph appears with coefficient $-(d-1)$ in the Bethe entropy function definition, we see that for a full-edge the corresponding coefficient must be $-(2-1) = -1$, and that for a half-edge the corresponding coefficient must be $-(1-1) = 0$.

Definition 15 For any $T \in \mathbb{R}_{>0}$, the Bethe partition function associated with some NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ is defined to be

$$Z_B \triangleq \exp \left(-\frac{1}{T} \cdot \min_{\beta \in \mathcal{B}} F_B(\beta) \right).$$

□

Let us comment on a variety of issues with respect to the above definition of the Bethe partition function.

- Note that here Z_B is *defined* such that a similar statement can be made as in Lemma 8.
- For NFGs *without* cycles one can show that $Z_B = Z_G$.
- The Bethe free energy function of NFGs *with* cycles can have non-global local extrema, which is in contrast to the Gibbs free energy function which is convex and therefore has no non-global local extrema.
- Similar to Lemma 10, we can consider a modified Bethe free energy function that equals the Bethe free energy function for β 's that are compatible with some $\mathbf{a}_{\text{half}} \in \mathcal{A}_{\text{half}}$, and that is infinite otherwise. Of course, the modified Bethe partition function will be a function of $\mathbf{a}_{\text{half}} \in \mathcal{A}_{\text{half}}$.

In the next section, we will present an alternative characterization of the local marginal polytope, namely in terms of so-called finite graph covers, thereby generalizing some observations that were made in [4], [5]. This will then lead the way to Section VII, where we will show that the Bethe entropy function, and consequently also the Bethe partition function, cannot only be characterized *analytically* (as was done here in Definitions 14 and 15), but also *combinatorially*. This combinatorial approach is based on counting valid configurations in finite graph covers of the underlying NFG.

VI. FINITE GRAPH COVERS

This section reviews the concept of a finite graph cover of a graph; for more details we refer the interested reader to [59]. We also refer to [4], [5], [60]–[62], where finite graph covers were used in the context of coding theory, especially for the analysis of linear programming decoding and message-passing iterative decoding.

Definition 16 (see, e.g., [59], [63]) A cover of a graph G with vertex set \mathcal{V} and edge set \mathcal{E} is a graph \tilde{G} with vertex set $\tilde{\mathcal{V}}$ and edge set $\tilde{\mathcal{E}}$, along with a surjection $\pi : \tilde{\mathcal{V}} \rightarrow \mathcal{V}$ which is a graph homomorphism (i.e., π takes adjacent vertices of G to adjacent vertices of \tilde{G}) such that for each vertex $v \in \mathcal{V}$ and each $\tilde{v} \in \pi^{-1}(v)$, the neighborhood $\partial(\tilde{v})$ of \tilde{v} is mapped bijectively to $\partial(v)$. A cover is called an M -cover, where $M \in \mathbb{Z}_{>0}$, if $|\pi^{-1}(v)| = M$ for every vertex v in \mathcal{V} .³ □

A consequence of this definition is that if \tilde{G} is an M -cover of G then we can choose its vertex set $\tilde{\mathcal{V}}$ to be $\tilde{\mathcal{V}} \triangleq \mathcal{V} \times [M]$: if $(v, m) \in \tilde{\mathcal{V}}$ then $\pi((v, m)) = v$ and if $((v_1, m_1), (v_2, m_2)) \in \tilde{\mathcal{E}}$ then $\pi(\{(v_1, m_1), (v_2, m_2)\}) = \{v_1, v_2\}$. Another consequence is that any M_2 -cover of any M_1 -cover of the base graph is an $(M_2 \cdot M_1)$ -cover of the base graph.

³The number M is also known as the degree of the cover. (Not to be confused with the degree of a vertex.)

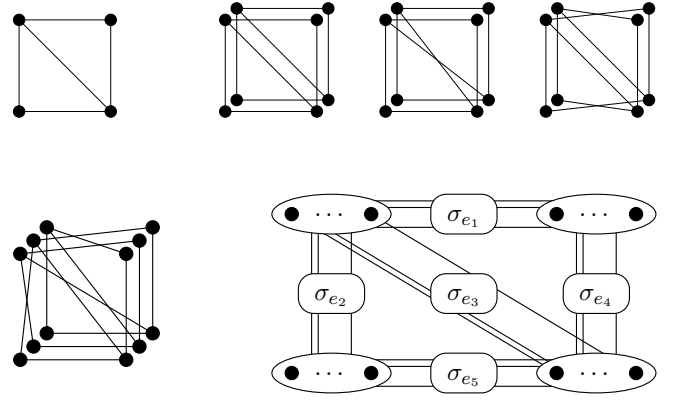


Fig. 5. Top left: base graph G . Top right: sample of possible 2-covers of G . Bottom left: a possible 3-cover of G . Bottom right: a possible M -cover of G . Here, $\sigma_{e_1}, \dots, \sigma_{e_5}$ are arbitrary edge permutations.

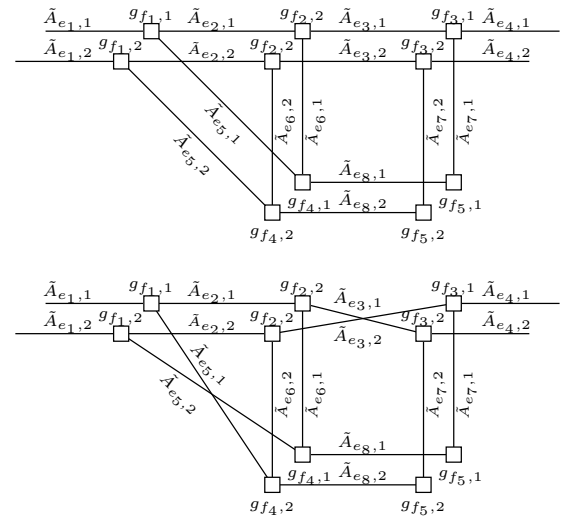


Fig. 6. Two possible 2-covers of the NFG N that is shown in Fig. 3.

Example 17 ([5]) Let G be a (base) graph with 4 vertices and 5 edges as shown in Fig. 5 (top left). Figs. 5 (top right), 5 (bottom left), and 5 (bottom right) show, respectively, possible 2-, 3-, and M -covers of G . Note that any 2-cover of G must have $8 = 2 \cdot 4$ vertices and $10 = 2 \cdot 5$ edges, that any 3-cover of G must have $12 = 3 \cdot 4$ vertices and $15 = 3 \cdot 5$ edges, and that any M -cover must have $M \cdot 4$ vertices and $M \cdot 5$ edges. As depicted in Fig. 5 (bottom right), any M -cover of G is entirely specified by $|\mathcal{E}|$ edge permutations, where \mathcal{E} is the edge set of G . □

As we can see from this example, an M -cover \tilde{G} of G may consist of several connected components also if G consists of only one connected component. In general, letting $\#G$ and $\#\tilde{G}$ denote the number of connected components of G and \tilde{G} , respectively, one can easily verify that

$$\#G \leq \#\tilde{G} \leq M \cdot \#G. \quad (14)$$

Because NFGs are graphs, we can also consider finite graph covers of NFGs, as is done in the next example. (Note that we do *not* apply edge permutations to copies of half-edges.)

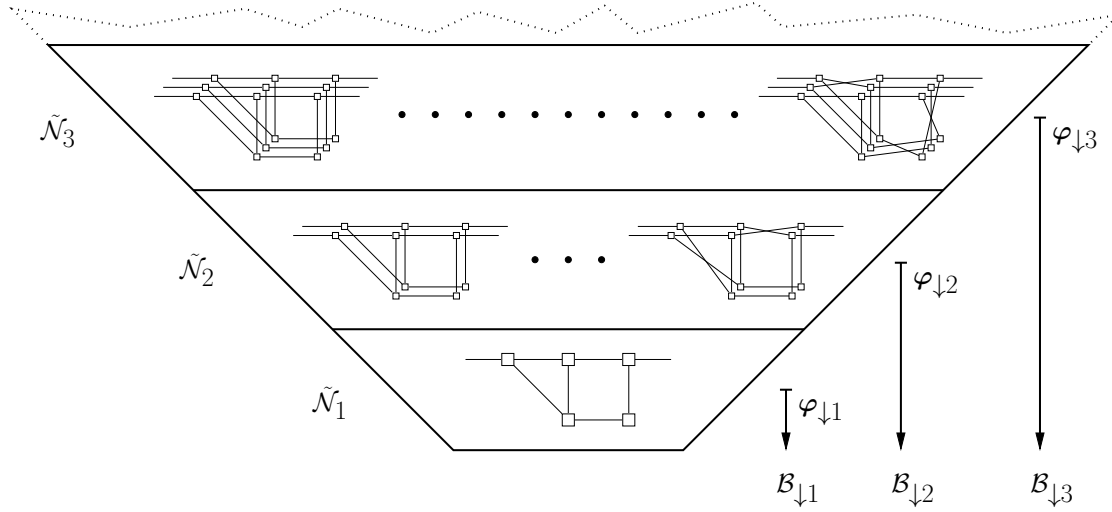


Fig. 7. Hierarchy of all finite graph covers of the (base) NFG N from Fig. 3, where the M th level lists all graphs in \tilde{N}_M , i.e., all M -covers. (For a base NFG with $|\mathcal{E}_{\text{full}}|$ full-edges, there are $|\tilde{N}_M| = (M!)^{|\mathcal{E}_{\text{full}}|}$ graph covers at the M th level.) The pseudo-marginal mappings φ_M , $M \in \mathbb{Z}_{>0}$, and their images are specified in Definitions 22 and 24.

Example 18 Consider again the NFG N that is discussed in Example 1 and depicted in Fig. 3. Two possible 2-covers of this (base) NFG are shown in Fig. 6. The first graph cover is “trivial” in the sense that it consists of two disjoint copies of the NFG in Fig. 3. The second graph cover is “more interesting” in the sense that the edge permutations are such that the two copies of the base NFG are intertwined. (Of course, both graph covers are equally valid.)

Note that the M copies of a function node g_f are denoted by $\{g_{f,m}\}_{m \in [M]}$, and that the M copies of a variable label A_e are denoted by $\{\tilde{A}_{e,m}\}_{m \in [M]}$. In that respect, we chose the variable labels to be such that if a full-edge e connects function nodes f_i and f_j , $i < j$, then the variable label $\tilde{A}_{e,m}$, $m \in [M]$, will be associated with the edge that connects the function nodes (f_i, m) and $(f_j, \sigma_e(m))$, where $\sigma_e : [M] \rightarrow [M]$ describes the permutation that is applied to the M copies of the edge e . Similarly, if a half-edge e is connected to the function node f then the variable label $\tilde{A}_{e,m}$, $m \in [M]$, will be associated with the half-edge that is connected to the function node (f, m) . That being said, it is important to note that the results that are presented in this paper are invariant to the chosen labeling convention. \square

Definition 19 Consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$. We define the set \tilde{N}_M to be the set of all M -covers \tilde{N} of N . \square

Note that in this definition we consider only labeled M -covers as follows:

- All vertices of an M -cover have distinct labels, say (f, m) or $g_{f,m}$ with $(f, m) \in \mathcal{F} \times [M]$.
- All edges of an M -cover have distinct labels, say (e, m) or $A_{e,m}$ with $(e, m) \in \mathcal{E} \times [M]$.
- We do not identify M -covers whose graphs are isomorphic but whose vertex labels are distinct.
- However, we do identify M -covers whose graphs (including vertex labels) are isomorphic but whose edge labels are distinct.

(For reasons of simplicity, these vertex and edge labels are sometimes omitted in drawings.) These conventions are reflected in the following lemma that counts M -covers.

Lemma 20 Consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$. Then

$$|\tilde{N}_M| = (M!)^{|\mathcal{E}_{\text{full}}(N)|}. \quad (15)$$

Proof: An M -cover \tilde{N} of N can be obtained as follows.

- 1) For every $f \in \mathcal{F}(N)$, draw M copies of f (with distinct labels).
- 2) For every $e = \{f, f'\} \in \mathcal{E}_{\text{full}}(N)$, connect the M copies of f and the M copies of f' by M disjoint edges.
- 3) For every $e = \{f\} \in \mathcal{E}_{\text{half}}(N)$, attach a half-edge to the M copies of f .

Because the second step can be done independently for every $e \in \mathcal{E}_{\text{full}}(N)$, because for every such edge there are $M!$ ways of connecting the M copies of f and the M copies of f' by M disjoint edges, and because the obtained M -covers are all distinct, the result follows. \blacksquare

Example 21 Consider again the NFG N that is discussed in Example 1 and depicted in Fig. 3. We can order all the finite graph covers of this base NFG according to the hierarchy shown in Fig. 7, where the M th level lists all graphs in \tilde{N}_M , i.e., all M -covers. (Note that there is exactly one 1-cover, namely the base NFG itself.) The inverted pyramid alludes to the fact that the number of M -covers is growing with M , i.e., $|\tilde{N}_M|$ is growing with M , see (15). \square

The following definition specifies a collection of mappings that will be crucial for the rest of this section and for the next section. These mappings are inspired by similar mappings that appear in the method of types, see Definition 11.

Definition 22 Let $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG with local marginal polytope \mathcal{B} (see Definition 13). For any $M \in \mathbb{Z}_{>0}$

we define the pseudo-marginal mapping

$$\varphi_M : \left\{ (\tilde{N}, \tilde{c}) \mid \tilde{N} \in \tilde{\mathcal{N}}_M, \tilde{c} \in \mathcal{C}(\tilde{N}) \right\} \rightarrow \mathcal{B},$$

$$(\tilde{N}, \tilde{c}) \mapsto \beta.$$

Here, for a given $\tilde{N} \in \tilde{\mathcal{N}}_M$ and $\tilde{c} \in \mathcal{C}(\tilde{N})$, the components of β are defined as follows

$$\beta_{f, \mathbf{a}_f} \triangleq \frac{1}{M} \sum_{m \in [M]} [\tilde{c}_{f,m} = \mathbf{a}_f], \quad f \in \mathcal{F}, \mathbf{a}_f \in \mathcal{A}_f, \quad (16)$$

$$\beta_{e, \mathbf{a}_e} \triangleq \frac{1}{M} \sum_{m \in [M]} [\tilde{c}_{e,m} = \mathbf{a}_e], \quad e \in \mathcal{E}, \mathbf{a}_e \in \mathcal{A}_e. \quad (17)$$

(In the above expressions we have used Iverson's convention that was defined in Section I-G.) \square

Note that one cannot mention a valid configuration \tilde{c} without mentioning the M -cover \tilde{N} in which it lives. Therefore, although the expressions in (16) and (17) do not involve the M -cover \tilde{N} explicitly, the domain of φ_M must be over (graph, valid configuration)-pairs (\tilde{N}, \tilde{c}) , and not just over valid configurations \tilde{c} .

Example 23 Consider again the NFG N that is discussed in Example 4 and depicted in Fig. 4, which goes back to Example 1 and Fig. 3. A possible 2-cover \tilde{N} of N is shown in Fig. 8. Applying the pseudo-marginal mapping to the valid configuration $\tilde{c} \in \mathcal{C}(\tilde{N})$ shown in Fig. 8, we obtain the pseudo-marginal vector $\beta \triangleq \varphi_M(\tilde{c})$ with the following components. (We show only a selection of the obtained pseudo-marginals.)

- For $\beta_{f_1, \mathbf{a}_1}$ with $\mathbf{a}_1 = (a_{e_1}, a_{e_2}, a_{e_3})$:

$$\beta_{f_1, (000)} = 1, \beta_{f_1, (001)} = 0, \beta_{f_1, (010)} = 0, \beta_{f_1, (011)} = 0,$$

$$\beta_{f_1, (100)} = 0, \beta_{f_1, (101)} = 0, \beta_{f_1, (110)} = 0, \beta_{f_1, (111)} = 0.$$

- For $\beta_{f_2, \mathbf{a}_2}$ with $\mathbf{a}_2 = (a_{e_2}, a_{e_3}, a_{e_6})$:

$$\beta_{f_2, (000)} = \frac{1}{2}, \beta_{f_2, (001)} = 0, \beta_{f_2, (010)} = 0, \beta_{f_2, (011)} = \frac{1}{2},$$

$$\beta_{f_2, (100)} = 0, \beta_{f_2, (101)} = 0, \beta_{f_2, (110)} = 0, \beta_{f_2, (111)} = 0.$$

- For $\beta_{f_5, \mathbf{a}_5}$ with $\mathbf{a}_5 = (a_{e_7}, a_{e_8})$:

$$\beta_{f_5, (00)} = \frac{1}{2}, \beta_{f_5, (01)} = 0, \beta_{f_5, (10)} = 0, \beta_{f_5, (11)} = \frac{1}{2}.$$

- For $\beta_{e_4, \mathbf{a}_4}$:

$$\beta_{e_4, 0} = 0, \beta_{e_4, 1} = 1.$$

- For $\beta_{e_7, \mathbf{a}_7}$:

$$\beta_{e_7, 0} = \frac{1}{2}, \beta_{e_7, 1} = \frac{1}{2}.$$

\square

The upcoming Theorem 25 will show that the local marginal polytope \mathcal{B} is a valid choice as a co-domain of the pseudo-marginal mapping φ_M . For this theorem, the following definition is useful.

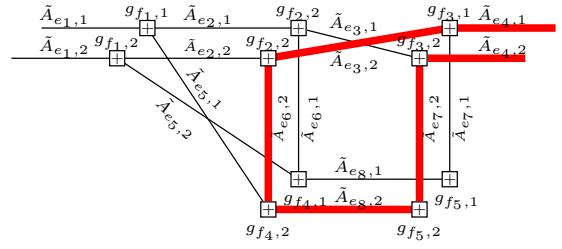


Fig. 8. Valid configuration \tilde{c} on a possible 2-cover of the NFG N in Fig. 3. (This 2-cover is identical to the second 2-cover in Fig. 6.) For every $(e, m) \in \mathcal{E} \times [M]$, if $\tilde{c}_{e,m} = 0$ then the edge (e, m) is thin and in black, whereas if $\tilde{c}_{e,m} = 1$ then the edge (e, m) is thick and in red. (See Example 23 for more details.)

Definition 24 Consider an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$. For every $M \in \mathbb{Z}_{>0}$, we define \mathcal{B}'_M to be the image of the pseudo-marginal mapping φ_M , i.e.,

$$\mathcal{B}'_M \triangleq \text{image}(\varphi_M).$$

Moreover, we define \mathcal{B}' to be the union of all \mathcal{B}'_M , i.e.,

$$\mathcal{B}' \triangleq \bigcup_{M \in \mathbb{Z}_{>0}} \mathcal{B}'_M. \quad (18)$$

\square

In words, \mathcal{B}' is the set where for every $\beta \in \mathcal{B}'$ there is some $M \in \mathbb{Z}_{>0}$ such that there is an M -cover $\tilde{N} \in \tilde{\mathcal{N}}_M$ with a valid configuration $\tilde{c} \in \mathcal{C}(\tilde{N})$ in it such that \tilde{c} maps down to β under the pseudo-marginal mapping φ_M . Generalizing the language of [61], we will call

- \mathcal{B}'_M : set of all M -cover lift-realizable ps.-marg. vectors,
- \mathcal{B}' : set of all lift-realizable pseudo-marginal vectors.

For any $M_1, M_2 \in \mathbb{Z}_{>0}$ with M_1 dividing M_2 , one can show the following chain of set inclusions

$$\mathcal{B}'_{M_1} \subseteq \mathcal{B}'_{M_2} \subseteq \mathcal{B}'.$$

The second set inclusion follows from (18); we leave it as an exercise for the reader to verify the first set inclusion.

Theorem 25 Let $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG with local marginal polytope \mathcal{B} . The set of all lift-realizable pseudo-marginal vectors satisfies

$$\mathcal{B}' = \mathcal{B} \cap \mathbb{Q}^{\dim(\mathcal{B})},$$

which implies that \mathcal{B}' is dense in \mathcal{B} . Moreover, all vertices of \mathcal{B} are in \mathcal{B}' .

Proof: This is a more or less a straightforward extension of the characterization in [4], [5] of the fundamental polytope in terms of valid configurations in finite graph covers. We omit the details. \blacksquare

Let us conclude this section with a few comments.

Remark 26 Let $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG with local marginal polytope \mathcal{B} . For any $M \in \mathbb{Z}_{>0}$, consider the set of all M -cover lift-realizable pseudo-marginal vectors \mathcal{B}'_M .

- It holds that

$$|\mathcal{B}'_M| \leq (M+1)^{\dim(\mathcal{B})}.$$

This follows from the observation that every component of $\beta \in \mathcal{B}'_M$ takes values in the set $\{\frac{0}{M}, \frac{1}{M}, \frac{2}{M}, \dots, \frac{M}{M}\}$. Because $\dim(\mathcal{B})$ is a fixed number for a given NFG \mathbf{N} , it follows that the number of elements of \mathcal{B}'_M grows at most polynomially in M . This important fact will allow us to use the method of types in the next section. (Compare this observation also with a similar statement in Lemma 12.)

- Although the focus of this paper is mostly on the behavior of \mathcal{B}'_M when M goes to infinity, the set \mathcal{B}'_M for $M = 1$, i.e., the set \mathcal{B}'_1 , is also of special interest. The reason for this is that $\text{conv}(\mathcal{B}'_1)$ contains all pseudo-marginal vectors that are globally realizable. Here, a pseudo-marginal vector β is called globally realizable [58] when there is a $\mathbf{p} \in \Pi_{\mathcal{C}}$ such that β contains the true marginals of \mathbf{p} , i.e.,

$$\begin{aligned} \beta_{f, \mathbf{a}_f} &= \sum_{\mathbf{c}: \mathbf{c}_f = \mathbf{a}_f} \mathbf{p}_{\mathbf{c}}, \quad f \in \mathcal{F}, \mathbf{a}_f \in \mathcal{A}_f, \\ \beta_{e, a_e} &= \sum_{\mathbf{c}: c_e = a_e} \mathbf{p}_{\mathbf{c}}, \quad e \in \mathcal{E}, a_e \in \mathcal{A}_e. \end{aligned}$$

- For the NFG \mathbf{N} discussed in Examples 4 and 23, one can verify that the local marginal polytope of \mathbf{N} satisfies

$$\mathcal{B} \supsetneq \text{conv}(\mathcal{B}'_1), \quad (19)$$

i.e., \mathcal{B} is strictly larger than $\text{conv}(\mathcal{B}'_1)$. This can be shown as follows. Consider the valid configuration $\tilde{\mathbf{c}}$ of the 2-cover shown in Fig. 8: its associated pseudo-marginal vector β does not lie in $\text{conv}(\mathcal{B}'_1)$. Indeed, because all variable alphabets are $\{0, 1\}$ and because $\beta_{e,0} = 1 - \beta_{e,1}$ for all $e \in \mathcal{E}$, one can verify that the condition that the vector β is in $\text{conv}(\mathcal{B}'_1)$ is equivalent to the condition that the vector

$$(\beta_{e_1,1}, \beta_{e_2,1}, \dots, \beta_{e_7,1}) = \left(\frac{0}{2}, \frac{0}{2}, \frac{1}{2}, \frac{2}{2}, \frac{0}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right)$$

is in the convex hull of the set \mathcal{C} of valid configurations of \mathbf{N} as listed in (3). However, the latter is not the case. Therefore, $\text{conv}(\mathcal{B}'_2) \supsetneq \text{conv}(\mathcal{B}'_1)$. Combining this with $\mathcal{B} \supseteq \text{conv}(\mathcal{B}'_2)$, we find that the \mathcal{B} satisfies (19). In conclusion, valid configurations like $\tilde{\mathbf{c}}$ in Fig. 8 are the reason why \mathcal{B} is strictly larger than $\text{conv}(\mathcal{B}'_1)$. \square

VII. COUNTING IN FINITE GRAPH COVERS

The definition of the Bethe entropy function and the Bethe partition function in Definitions 14 and 15, respectively, were entirely analytical. In this subsection we will present a combinatorial characterization of these functions in terms of counting certain valid configurations in graph covers, a characterization that was first outlined in [64], [65].

We start with the definition of a certain averaging operator. This definition is motivated by the fact that many results in this section are based on associating a real number to every M -cover of a base NFG and on computing the average of this value over all M -covers.

Definition 27 Let $\mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG. For any $M \in \mathbb{Z}_{>0}$ and any function $\chi_M : \tilde{\mathcal{N}}_M \rightarrow \mathbb{R}$ we define the averaging operator to be

$$\langle \chi_M(\tilde{\mathbf{N}}) \rangle_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \triangleq \frac{1}{|\tilde{\mathcal{N}}_M|} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \chi_M(\tilde{\mathbf{N}}). \quad \square$$

A. The Bethe Entropy Function

The next definition introduces the function that will be key towards the promised combinatorial characterization of the Bethe entropy function.

Definition 28 Let $\mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG and let \mathcal{B}' be its set of all lift-realizable pseudo-marginal vectors. Then, for every $M \in \mathbb{Z}_{>0}$ and every $\beta \in \mathcal{B}'$ we define

$$\begin{aligned} \chi_{M, \beta} : \tilde{\mathcal{N}}_M &\rightarrow \mathbb{R}, \\ \tilde{\mathbf{N}} &\mapsto \left| \left\{ \tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}}) \mid \varphi_M(\tilde{\mathbf{N}}, \tilde{\mathbf{c}}) = \beta \right\} \right|. \end{aligned} \quad (20)$$

\square

Note that for an M -cover $\tilde{\mathbf{N}}$ of \mathbf{N} the value of $\chi_{M, \beta}(\tilde{\mathbf{N}})$ represents the number of valid configurations in $\tilde{\mathbf{N}}$ that map down to β . Consequently,

$$\bar{C}_M(\beta) \triangleq \langle \chi_{M, \beta}(\tilde{\mathbf{N}}) \rangle_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \quad (21)$$

is the average number of valid configurations that map down to β , where the averaging is over all M -covers of \mathbf{N} . (Observe that this is the same $\bar{C}_M(\beta)$ as in Section I.) Letting φ_M^{-1} denote the inverse of the mapping φ_M , the quantity $\bar{C}_M(\beta)$ can also be written in terms of the pre-image of β under the mapping φ_M , i.e.,

$$\bar{C}_M(\beta) = \frac{|\varphi_M^{-1}(\beta)|}{|\tilde{\mathcal{N}}_M|}.$$

Lemma 29 Let $\mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be some NFG, and for every $M \in \mathbb{Z}_{>0}$ let \mathcal{B}'_M be its set of all M -cover lift-realizable pseudo-marginal vectors. Then for every $\beta \in \mathcal{B}'_M$ we have

$$\bar{C}_M(\beta) = \left(\prod_{f \in \mathcal{F}} \binom{M}{M \cdot \beta_f} \right) \cdot \left(\prod_{e \in \mathcal{E}_{\text{full}}} \binom{M}{M \cdot \beta_e} \right)^{-1},$$

where we have used the multinomial coefficients

$$\binom{M}{M \cdot \beta_f} \triangleq \frac{M!}{\prod_{\mathbf{a}_f} (M \beta_{f, \mathbf{a}_f})!}, \quad (22)$$

$$\binom{M}{M \cdot \beta_e} \triangleq \frac{M!}{\prod_{a_e} (M \beta_{e, a_e})!}. \quad (23)$$

(Note that the components of $M \cdot \beta$ are non-negative integers and so these expressions are indeed well defined.)

Proof: See Appendix A. \blacksquare

Note that the multinomial coefficients that appear in the above expression for $\bar{C}_M(\beta)$ have different origins. Namely, the multinomial coefficients in the numerator of $\bar{C}_M(\beta)$ stem from counting locally valid configurations at the function

nodes (see the proof of Lemma 29 in Appendix A for the definition of “locally valid configurations”), whereas the multinomial coefficients in the denominator of $\bar{C}_M(\beta)$ stem from counting the number of edge connections that lead to overall valid configurations, and from the division by the total number of M -covers.

The next theorem states the first main result of this section (and also of this paper). It connects the asymptotic behavior of $\bar{C}_M(\beta)$ with the Bethe entropy function value of β . Therefore, this result gives the promised combinatorial characterization of the Bethe entropy function value of β . (The second main result of this section will be the combinatorial characterization of the Bethe partition function presented in Theorem 33.)

Theorem 30 *Let $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be some NFG and let \mathcal{B}' be its set of lift-realizable pseudo-marginal vectors. For any $\beta \in \mathcal{B}'$ we have*

$$\limsup_{M \rightarrow \infty} \frac{1}{M} \log (\bar{C}_M(\beta)) = H_B(\beta).$$

Proof: There are infinitely many $M \in \mathbb{Z}_{>0}$ such that $\beta \in \mathcal{B}'_M$. This can be seen as follows. Namely, by definition of \mathcal{B}' , there must be at least one $M^* \in \mathbb{Z}_{>0}$ such that $\beta \in \mathcal{B}'_{M^*}$. However, because $\beta \in \mathcal{B}'_{M^*}$ implies that $\beta \in \mathcal{B}'_M$ holds for any $M \in \mathbb{Z}_{>0}$ that is divisible by M^* , there are in fact infinitely many $M \in \mathbb{Z}_{>0}$ such that $\beta \in \mathcal{B}'_M$. The theorem statement then follows by combining Lemma 29, the results

$$\begin{aligned} \binom{M}{M \cdot \beta_f} &= \exp \left(-M \cdot \sum_{\mathbf{a}_f} \beta_{f, \mathbf{a}_f} \log (\beta_{f, \mathbf{a}_f}) + o(M) \right), \\ \binom{M}{M \cdot \beta_e} &= \exp \left(-M \cdot \sum_{\mathbf{a}_e} \beta_{e, \mathbf{a}_e} \log (\beta_{e, \mathbf{a}_e}) + o(M) \right), \end{aligned}$$

for $\beta \in \mathcal{B}'_M$ (which are consequences of Stirling’s approximation of the factorial function), and the definition of the Bethe entropy function in Definition 14. ■

A straightforward consequence of Theorem 30 is that for $\beta \in \mathcal{B}'_M$ we have

$$\bar{C}_M(\beta) = \exp (M \cdot H_B(\beta) + o(M)). \quad (24)$$

Therefore, the Bethe entropy function value of β has the meaning of being the asymptotic growth rate of the average number of valid configurations in M -covers that map down to β , where the averaging is over all M -covers of N , and where asymptotic is in the sense that M goes to infinity.

At this point, we encourage the reader to compare the observations that were made so far in this section with similar statements that were made in Lemma 12 with respect to $C_M(\mathbf{q})$. There are many similarities, but also some key differences. One key difference is the following:

- In the setup of the present section, we count the average number of certain valid configurations in M -covers of some NFG N . Most importantly, every M -cover is obtained by suitably “intertwining” M independent and identical copies of N .
- In the setup of Section IV, we count certain valid configurations in N^M , which corresponds to counting certain

valid configurations in the M -cover of N that consists of M independent and identical copies of N .⁴

In conclusion: when counting certain valid configurations in “intertwined” M -covers we get the Bethe entropy function, whereas when counting certain valid configurations in “non-intertwined” M -covers we get the Gibbs entropy function. (See also the comments at the end of the upcoming Section VII-D.)

B. The Bethe Average Energy Function

In this subsection we show how the global function of an M -cover of some base NFG can be expressed in terms of the Bethe average energy function of this base NFG.

Theorem 31 *Let $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be some NFG and let $M \in \mathbb{Z}_{>0}$. Then for any M -cover \tilde{N} of N and any $\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{N})$ we have*

$$-\frac{1}{M} \log g_{\tilde{N}}(\tilde{\mathbf{c}}) = U_B(\beta) \Big|_{\beta = \varphi_M(\tilde{N}, \tilde{\mathbf{c}})}.$$

(Note that this expression does not involve a limit $M \rightarrow \infty$.)

Proof: Let $\tilde{\mathcal{F}}$ be the set of function nodes of \tilde{N} . Then

$$g_{\tilde{N}}(\tilde{\mathbf{c}}) = \prod_{\tilde{f} \in \tilde{\mathcal{F}}} g_{\tilde{f}}(\tilde{\mathbf{c}}_{\tilde{f}}) = \prod_{f \in \mathcal{F}} \prod_{\mathbf{a}_f} (g_f(\mathbf{a}_f))^{M \cdot \beta_{f, \mathbf{a}_f}}.$$

Taking the logarithm on both sides, multiplying both sides by $-1/M$, and using the definition of the Bethe average energy function (see Definition 14), we obtain the expression stated in the theorem. ■

Recall the definition of globally realizable pseudo-marginal vectors from Remark 26. One can easily show that for every $\beta \in \text{conv}(\mathcal{B}'_1)$ it holds that

$$U_B(\beta) = U_G(\mathbf{p}),$$

where $\mathbf{p} \in \Pi_{\mathcal{C}}$ is the distribution whose marginals are given by β . In this sense, the Bethe average energy function can be seen as a “straightforward extension” of the Gibbs average energy function from the domain $\text{conv}(\mathcal{B}'_1)$ to the domain \mathcal{B} .

C. The Bethe Free Energy Function

An immediate consequence of Theorems 30 and 31 is that for any temperature $T \in \mathbb{R}_{>0}$, any $M \in \mathbb{Z}_{>0}$, any M -cover \tilde{N} of N , and any $\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{N})$ it holds that

$$g_{\tilde{N}}(\tilde{\mathbf{c}})^{1/T} \cdot \bar{C}_M(\beta) = \exp \left(-\frac{M}{T} \cdot F_B(\beta) + o(M) \right), \quad (25)$$

where $\beta \triangleq \varphi_M(\tilde{N}, \tilde{\mathbf{c}})$.

⁴More precise would be “...which corresponds to counting certain valid configurations in one of the M -covers of N that consists of M independent and identical copies of N .”

D. The Degree- M Bethe Partition Function

The above developments motivate the following definition of a degree- M Bethe partition function, which, as we will show, has the property that in the limit $M \rightarrow \infty$ it converges to the Bethe partition function. Note that in contrast to the definition of the Bethe partition function in Definition 15, which was analytical, the definition of the degree- M Bethe partition function is combinatorial.

Definition 32 Let $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG. For any temperature $T \in \mathbb{R}_{>0}$ and any $M \in \mathbb{Z}_{>0}$, we define the degree- M Bethe partition function to be

$$Z_{B,M}(N) \triangleq \sqrt[M]{\left\langle Z_G(\tilde{N}) \right\rangle_{\tilde{N} \in \tilde{\mathcal{N}}_M}}.$$

(Note that the right-hand side of the above expression is based on the Gibbs partition function, see (4), and not on the Bethe partition function.) \square

From the above expression we see that the degree- M Bethe partition function is defined to be the M th root of the average Gibbs partition function, where the averaging is done over all M -covers of N .

With this we are in a position to formulate the second main result of this section (and of this paper).

Theorem 33 For any NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ and any temperature $T \in \mathbb{R}_{>0}$ it holds that

$$\limsup_{M \rightarrow \infty} Z_{B,M}(N) = Z_B(N).$$

Proof: See Appendix B. \blacksquare

Example 34 For improving one's understanding of $Z_{B,M}$, it is helpful to explicitly compute this quantity for small NFGs and small values of M . To this end, consider the NFG N in the lower left corner of Fig. 9. Assume that the variable alphabets and local functions are defined analogously to variable alphabets and local functions of the NFG in Example 4. (The same NFG was also discussed in [5, Example 29] and in [60, Example 2.4].)

One can easily verify that all valid configurations take on the global function value one. Moreover, because N does not have half-edges, the set $\{e \in \mathcal{E} \mid c_e = 1\}$ associated with a valid configuration c forms a cycle or an edge-disjoint union of cycles in N . With this, the set $\mathcal{C}(N)$ of valid configurations contains four elements, as shown in the last row of Fig. 9, i.e.,

$$Z_G(N) = 4.$$

Because N has seven edges, there are $2^7 = 128$ distinct 2-covers: 32 of them are (when omitting the cover-related parts of the vertex and edge labels) isomorphic to $\tilde{N}_{2,1}$, 32 of them are isomorphic to $\tilde{N}_{2,2}$, 32 of them are isomorphic to $\tilde{N}_{2,3}$, and 32 of them are isomorphic to $\tilde{N}_{2,4}$ shown on the left-hand side of Fig. 9. One can verify that

$$Z_G(\tilde{N}_{2,1}) = 16, \quad Z_G(\tilde{N}_{2,2}) = Z_G(\tilde{N}_{2,3}) = Z_G(\tilde{N}_{2,4}) = 8.$$

Therefore, the degree-2 Bethe partition function is

$$Z_{B,2}(N) = \sqrt[2]{\frac{1}{128} \sum_{h \in [4]} 32 Z_G(\tilde{N}_{2,h})} = \sqrt[2]{10} = 3.162 \dots$$

Some comments:

- The 2-cover $\tilde{N}_{2,1}$ consists of two copies of N and consequently we have $Z_G(\tilde{N}_{2,1}) = (Z_G(N))^2 = 4^2 = 16$. If $Z_G(\tilde{N}) = (Z_G(N))^2$ were true for all 2-covers of N , then $Z_{B,2}(N) = Z_G(N)$.
- As we can see from Fig. 9, there are 2-covers \tilde{N} such that $Z_G(\tilde{N}) \neq (Z_G(N))^2$. Therefore, it is not surprising that $Z_{B,2}(N) \neq Z_G(N)$.
- If all 2-covers were like $\tilde{N}_{2,h}$, $h \in [3]$, then the set of 2-cover lift-realizable pseudo-marginal vectors would satisfy $\text{conv}(\mathcal{B}'_2) = \text{conv}(\mathcal{B}'_1)$. However, the 2-cover $\tilde{N}_{2,4}$ contains some configurations whose associated pseudo-marginal vector does not lie within $\text{conv}(\mathcal{B}'_1)$. Therefore, $\text{conv}(\mathcal{B}'_2) \supsetneq \text{conv}(\mathcal{B}'_1)$, and so, because $\mathcal{B} \supsetneq \text{conv}(\mathcal{B}'_2)$, we literally see why the NFG N is an example where the local marginal polytope satisfies $\mathcal{B} \supsetneq \text{conv}(\mathcal{B}'_1)$. (See Remark 26 for a related observation.)
- As mentioned in Section I-E, one can also give a combinatorial characterization of the Bethe partition function of an NFG N in terms of computation trees and the universal cover \hat{N} of N . However, in many respects, finite graph covers are easier to deal with, and, as this example shows, many effects that are responsible for the similarities and differences between $Z_B(N)$ and $Z_G(N)$ are already visible in finite graph covers with small cover degree M . \square

As the following lemma shows, it is no coincidence that $Z_{B,2}(N)$ is a lower bound of $Z_G(N)$ for the NFG N in Example 35. Let $\#N$ be the number of connected components of N , when N is considered as a graph.

Lemma 35 Consider an NFG N as defined in Example 34, in particular without half-edges. For any $M \in \mathbb{Z}_{>0}$ it holds that

$$2^{-((M-1)/M) \cdot \#N} \cdot Z_G(N) \leq Z_{B,M}(N) \leq Z_G(N), \\ 2^{-\#N} \cdot Z_G(N) \leq Z_B(N) \leq Z_G(N). \quad (26)$$

Equivalently,

$$Z_{B,M}(N) \leq Z_G(N) \leq 2^{((M-1)/M) \cdot \#N} \cdot Z_{B,M}(N), \\ Z_B(N) \leq Z_G(N) \leq 2^{\#N} \cdot Z_B(N). \quad (27)$$

Proof: Because $Z_G(N)$ equals the number of cycles and edge-disjoint unions of cycles of N , we get

$$Z_G(N) = 2^{\text{circ}(N)}, \quad (28)$$

where

$$\text{circ}(N) = |\mathcal{E}(N)| - |\mathcal{F}(N)| + \#N \quad (29)$$

is the circuit rank of N . Similarly, for any M -cover \tilde{N} of N we obtain

$$Z_G(\tilde{N}) = 2^{\text{circ}(\tilde{N})}, \quad (30)$$

	NFG	valid configurations
2-cover $\tilde{N}_{2,4}$		
2-cover $\tilde{N}_{2,3}$		
2-cover $\tilde{N}_{2,2}$		
2-cover $\tilde{N}_{2,1}$		
1-cover N		

Fig. 9. NFGs that are used in Example 34, along with their valid configurations. Concerning the NFGs that appear in the “valid configurations” column: for every $(e, m) \in \mathcal{E} \times [M]$, if $\tilde{c}_{e,m} = 0$ then the edge (e, m) is thin and in black, whereas if $\tilde{c}_{e,m} = 1$ then the edge (e, m) is thick and in red.

where

$$\text{circ}(\tilde{N}) = |\mathcal{E}(\tilde{N})| - |\mathcal{F}(\tilde{N})| + \#\tilde{N}. \quad (31)$$

From straightforward graph-theoretic considerations of M -covers, in particular (14), it follows that

$$|\mathcal{E}(\tilde{N})| = M \cdot |\mathcal{E}(N)|, \quad (32)$$

$$|\mathcal{F}(\tilde{N})| = M \cdot |\mathcal{F}(N)|, \quad (33)$$

$$\#\tilde{N} \leq \#N \leq M \cdot \#\tilde{N}. \quad (34)$$

Combining (29), (31), and (32)–(34), we obtain

$$M \cdot \text{circ}(N) - (M-1) \cdot \#N \leq \text{circ}(\tilde{N}) \leq M \cdot \text{circ}(N).$$

Then, with the help of (28) and (30), we get

$$2^{-(M-1) \cdot \#N} \cdot (Z_G(N))^M \leq Z_G(\tilde{N}) \leq (Z_G(N))^M.$$

Plugging these expressions into the definition of $Z_{B,M}(N)$, see Definition 32, yields the result that was promised in the lemma statement. ■

We conclude this subsection with a few comments and observations.

- Of course, the complexity of computing $\#N$ and $\text{circ}(N)$ is polynomial in $|\mathcal{F}(N)|$ and $|\mathcal{E}(N)|$; therefore, the complexity of computing $Z_G(N)$ in Lemma 35 is polynomial in $|\mathcal{F}(N)|$ and $|\mathcal{E}(N)|$. Moreover, computing the lower and upper bounds in (27) is equally complex. The relevance of Example 34 and Lemma 35 is therefore not that an intractable partition function $Z_G(N)$ is approximated by lower and upper bounds based on a tractable Bethe partition function $Z_B(N)$, but to exhibit an example where explicit computations can easily be done and insight be gained into the formalism presented in this section.
- The inequalities in (26) and (27) can also easily be obtained by minimizing the Bethe free energy function and using the expression in Definition 15. However, using the combinatorial characterization of $Z_B(N)$ gives us additional insights why (26) holds. In fact, analyzing the proof of Lemma 35 we see that these inequalities are a straightforward consequence of the graph-theoretic inequalities

$$\#N \leq \#\tilde{N} \leq M \cdot \#N$$

that hold for any M -cover \tilde{N} of N , see also (14).

- Note that proving the inequality $Z_G(\tilde{N}) \leq (Z_G(N))^M$ for every M -cover \tilde{N} of N , as was done in the proof of Lemma 35, was also at the heart of the recent proof by Ruozzi [45] of the inequality $Z_B(N) \leq Z_G(N)$ for log-supermodular graphical models N . (This verified a conjecture by Sudderth, Wainwright, and Willsky [46].) Moreover, the paper [41] presented setups where the inequality $Z_G(\tilde{N}) \leq (Z_G(N))^M$ holds for every M -cover \tilde{N} of an NFG N whose partition function represents the permanent of a non-negative matrix, and pointed out setups where this inequality is conjectured to hold. Further NFGs where this inequality is conjectured to hold were listed by Watanabe [44].

$$\begin{array}{c} Z_{B,M}(N) \big|_{M \rightarrow \infty} = Z_B(N) \\ \downarrow \\ Z_{B,M}(N) \\ \downarrow \\ Z_{B,M}(N) \big|_{M=1} = Z_G(N) \end{array}$$

Fig. 10. The degree- M Bethe partition function of the NFG N for different values of M .

- When considering the value of $Z_{B,M}(N)$ from $M = 1$ to $M = \infty$, one goes from $Z_G(N)$ to $Z_B(N)$, see Fig. 10. It is worthwhile to consider the inequalities that appear in Lemma 35 under this perspective.
- We can write the ratio $Z_B(N)/Z_G(N)$ as the following telescoping product

$$\frac{Z_B(N)}{Z_G(N)} = \lim_{M \rightarrow \infty} \prod_{M' \in [M]} \frac{Z_{B,M'+1}(N)}{Z_{B,M'}(N)}.$$

Towards a better understanding of the ratio $Z_B(N)/Z_G(N)$, it might therefore be worthwhile to study the ratios $Z_{B,M+1}(N)/Z_{B,M}(N)$, $M \in \mathbb{Z}_{>0}$. We leave it as an open problem to see if general statements can be made about them.

- Let \tilde{N}_M^{\parallel} be the subset of \tilde{N}_M that contains all M -covers \tilde{N} that consist of M disconnected copies of N . It holds that $Z_G(\tilde{N}) = (Z_G(N))^M$ for any $\tilde{N} \in \tilde{N}_M^{\parallel}$, and so, trivially,

$$Z_G(N) \triangleq \sqrt[M]{\langle Z_G(\tilde{N}) \rangle_{\tilde{N} \in \tilde{N}_M^{\parallel}}}.$$

- If N does *not* contain any cycles, then

$$\tilde{N}_M^{\parallel} = \tilde{N}_M, \quad M \in \mathbb{Z}_{>0},$$

and so $Z_G(N) = Z_B(N)$.

- If N does contain cycles then

$$\tilde{N}_M^{\parallel} \subsetneq \tilde{N}_M, \quad M \geq 2.$$

Because usually $Z_G(\tilde{N}) \neq (Z_G(N))^M$ for $\tilde{N} \in \tilde{N}_M \setminus \tilde{N}_M^{\parallel}$, it is not surprising that usually $Z_G(N) \neq Z_B(N)$ for an NFG N with cycles.

E. Similarities and Differences w.r.t. the Replica Method

In this subsection we discuss similarities and differences between, on the one hand, the concepts and the mathematical expressions that have so far appeared in this section, and, on the other hand, concepts and mathematical expressions that appear in the replica theory (see, e.g., [34], [35, Chapter 8], [36, Appendix I]).

Let $N_{N,R}$ be an NFGs with “size parameter” N whose local functions depend on a random variable (or random vector) R . Assume that $N_{N,R}$ represents some physical system. Many interesting physical quantities about this physical system can then be derived from the normalized log-partition

function $\frac{1}{N} \log(Z_G(N_{N,R}))$. However, because this expression is usually not tractable, one studies the ensemble average $E[\frac{1}{N} \log(Z_G(N_{N,R}))]$, where the expectation value is w.r.t. R . If “measure concentration” happens for large N , then the normalized log-partition function of the “typical” NFG will be close to this expression for large N .

Direct evaluation of $E[\frac{1}{N} \log(Z_G(N_{N,R}))]$ is often not possible. Here is the point where the replica method comes in. Inspired by the equation

$$\log(z) = \lim_{M \downarrow 0} \frac{z^M - 1}{M}, \quad z \in \mathbb{R}_{>0},$$

where M is considered to be a real number, the replica method proposes the following reformulation of the above expectation value

$$E\left[\frac{1}{N} \log(Z_G(N_{N,R}))\right] = \lim_{M \downarrow 0} \frac{E[(Z_G(N_{N,R}))^M] - 1}{NM}.$$

One then notices that for *positive integers* M the term $(Z_G(N_{N,R}))^M$, which appears on the right-hand side of the above expression, corresponds to considering the partition function of M independent copies of $N_{N,R}$ (hence the name “replica theory”). After evaluating $E[(Z_G(N_{N,R}))^M]$ for positive integers M , one then *drops this requirement on M* , and evaluates the limit $M \downarrow 0$. This is the gist behind the replica method. Much more can, and needs to be said, for which we refer to [34], [35, Chapter 8], [36, Appendix I] (and references therein).

Clearly, there are similarities between the replica method and the developments in this paper. However, there are also some stark differences.

- The NFG $N_{N,R}$ depends on a random variable R , whereas the NFG N that is studied in this paper is deterministic. (Of course, our setup also allows NFGs that depend on a random variable, but that is not necessary.)
- Because the random variable R is the same in all M copies, there is the appearance of some “coupling effect” between the M copies of $N_{N,R}$ when evaluating $E[(Z_G(N_{N,R}))^M]$. This is in contrast to the “coupling effect” that appears in an M -cover of N as a result of the graph-cover construction process where the edges of M independent copies of N are permuted.
- The replica method is based on studying the limit $M \downarrow 0$, whereas this paper typically studies the limit $M \rightarrow \infty$. Moreover, this paper never drops the requirement that M is a positive integer.

Note that in coding theory, when studying the growth rate of the average Hamming weight enumerator of a code ensemble [66], one usually evaluates an expression like $\frac{1}{N} \log(E[Z_G(N_{N,R})])$. This quantity can be the same as the above-mentioned $E[\frac{1}{N} \log(Z_G(N_{N,R}))]$ but in general one can only state that

$$\frac{1}{N} \log(E[Z_G(N_{N,R})]) \geq E\left[\frac{1}{N} \log(Z_G(N_{N,R}))\right],$$

which is a consequence of Jensen’s inequality. For more information on these types of issues we refer to, e.g., [67].

Let us conclude this subsection by mentioning a recent paper by Mori [37] that was inspired by an earlier version of the present paper and that offers an alternative (and simpler) approach to some computations that are done in the context of the replica method. For more details we refer to Mori’s paper. See also [38].

VIII. NFGS FOR CHANNEL CODING

The main purpose of this section is to introduce some notation and concepts that will be useful for the next two sections, namely for Section IX on graph-cover decoding and for Section X on a connection between the minimum Hamming distance of a code and the non-concavity of the Bethe entropy function of some graphical model that represents this code.

For the following definition, we remind the reader of the sets \mathcal{C} and $\mathcal{C}_{\text{half}}$ that were specified in Definition 3 and the modified Gibbs partition function that was specified in Lemma 10.

Definition 36 Let \mathcal{X} be some finite set and let \mathcal{C}_{ch} be a length- n channel code over \mathcal{X} , i.e., $\mathcal{C}_{\text{ch}} \subseteq \mathcal{X}^n$. We say that an NFG $N(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ represents the code \mathcal{C}_{ch} if the following four conditions are satisfied.

- All local functions are indicator functions.
- For every $e \in \mathcal{E}_{\text{half}}$ we have $\mathcal{A}_e = \mathcal{X}$.
- The code \mathcal{C}_{ch} is the projection of \mathcal{C} to $\mathcal{E}_{\text{half}}$, i.e.,

$$\mathcal{C}_{\text{ch}} = \mathcal{C}_{\text{half}} \triangleq \{(c_e)_{e \in \mathcal{E}_{\text{half}}} \mid c \in \mathcal{C}\}.$$

- There is a $t_N \in \mathbb{Z}_{>0}$ such that

$$\left| \{c \in \mathcal{C} \mid (c_e)_{e \in \mathcal{E}_{\text{half}}} = x\} \right| = t_N \quad (\text{for all } x \in \mathcal{C}_{\text{half}}),$$

i.e., for every $x \in \mathcal{C}_{\text{half}}$ there are t_N valid configurations $c \in \mathcal{C}$ whose restriction to $\mathcal{E}_{\text{half}}$ equals x . \square

Let us comment on this definition.

- One can verify that the last condition is always satisfied in the following important special case: namely the case where all edge alphabets are equal to some group and all local functions represent indicator functions of subgroups of this group. (The proof of this statement uses the fact that all cosets of a subgroup have the same size. We leave the details to the reader.)
- Note that in Example 4, for every $x \in \mathcal{C}_{\text{half}}$, there are $t_N = 4$ valid configurations in \mathcal{C} whose restriction to $\mathcal{E}_{\text{half}}$ equals x .
- Usually, an NFG that represents a code is set up such that $t_N = 1$. However, sometimes it is more natural to set up N such that $t_N > 1$. For a more detailed discussion of this and related issues we refer the interested reader to, e.g., [68].

Example 37 Consider the length-10 code \mathcal{C}_{ch} over \mathbb{F}_2 defined by the parity-check matrix

$$H \triangleq \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix},$$

i.e., $\mathcal{C}_{\text{ch}} \triangleq \{\mathbf{x} \in \mathbb{F}_2^n \mid \mathbf{H} \cdot \mathbf{x}^\top = \mathbf{0}^\top \text{ (in } \mathbb{F}_2)\}$, where vectors are row vectors and where $(\cdot)^\top$ denotes vector transposition. This code can be represented by the NFG shown in Fig. 11 (left). Here, all edge alphabets are equal to \mathbb{F}_2 , all function nodes on the left-hand side represent indicator function nodes of repetition codes, and all function nodes on the right-hand side represent indicator function nodes of single parity-check codes. It can easily be verified that for this NFG we have $t_N = 1$. \square

This example is formalized in the following definition.

Definition 38 Consider a code \mathcal{C}_{ch} over \mathbb{F}_2 defined by some parity-check matrix $\mathbf{H} = [h_{j,i}]_{j \in \mathcal{J}, i \in \mathcal{I}}$, where \mathcal{J} and \mathcal{I} are the set of row and column indices of \mathbf{H} , respectively. The code \mathcal{C}_{ch} can be represented by an NFG $\mathbf{N}(\mathbf{H}) \triangleq \mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ as follows.

- The set of local function nodes is $\mathcal{F} \triangleq \mathcal{I} \cup \mathcal{J}$.
- The set of edges is $\mathcal{E} = \mathcal{E}_{\text{half}} \cup \mathcal{E}_{\text{full}}$, where $\mathcal{E}_{\text{half}} = \mathcal{I}$ and where $\mathcal{E}_{\text{full}} = \{(i, j) \in \mathcal{I} \times \mathcal{J} \mid h_{j,i} = 1\}$.
- For every $e \in \mathcal{E}$, the edge alphabet is $\mathcal{A}_e = \mathbb{F}_2$.
- For every $i \in \mathcal{I}$, the local function g_i equals the indicator function of a length- $(|\mathcal{E}_i|+1)$ repetition code.
- For every $j \in \mathcal{J}$, the local function g_j equals the indicator function of a length- $|\mathcal{E}_j|$ single parity-check code.

If the parity-check matrix \mathbf{H} is such that all columns of \mathbf{H} have Hamming weight d_L and all rows of \mathbf{H} have Hamming weight d_R , then \mathbf{H} is called a (d_L, d_R) -regular parity-check matrix. (For example, the parity-check matrix \mathbf{H} in Example 37 is $(3, 6)$ -regular.) If the parity-check matrix \mathbf{H} is sparsely populated then the code \mathcal{C}_{ch} is called a low-density parity-check (LDPC) code. Consequently, if the parity-check matrix \mathbf{H} of an LDPC code is (d_L, d_R) -regular then \mathcal{C}_{ch} is called a (d_L, d_R) -regular LDPC code, otherwise \mathcal{C}_{ch} is called an irregular LDPC code. \square

The following definition is a generalization of the definition of the fundamental polytope and the fundamental cone in [4], [5].

Definition 39 Let \mathcal{C}_{ch} be a code over \mathbb{F}_2 , let $\mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG that represents \mathcal{C}_{ch} , and let \mathcal{B} be the local marginal polytope of \mathbf{N} . We define the fundamental polytope \mathcal{P} and the fundamental cone \mathcal{K} to be, respectively,

$$\mathcal{P} \triangleq \{(\beta_{e,1})_{e \in \mathcal{E}_{\text{half}}} \mid \beta \in \mathcal{B}\},$$

$$\mathcal{K} \triangleq \text{conic}(\mathcal{P}).$$

Elements of \mathcal{P} and \mathcal{K} are called pseudo-codewords. \square

It can easily be verified that $\text{conv}(\mathcal{C}_{\text{ch}}) \subseteq \mathcal{P}$, i.e., that the fundamental polytope is a relaxation of the convex hull of the set of codewords. (Here the codewords are assumed to be embedded in \mathbb{R}^n , where n is the length of the code.)

The following definition is taken from [69].

Definition 40 Let \mathcal{C}_{ch} be a code over \mathbb{F}_2 , let $\mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ be an NFG that represents \mathcal{C}_{ch} , and let \mathcal{B} be the local marginal polytope of \mathbf{N} .

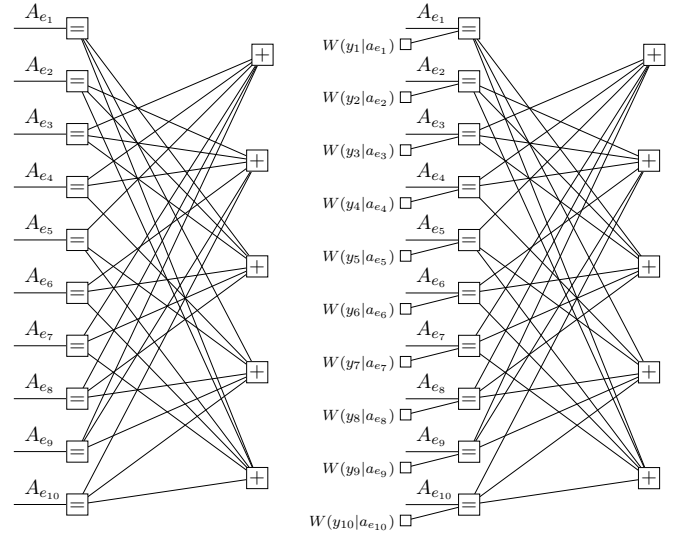


Fig. 11. Left: NFG representing the code \mathcal{C}_{ch} in Example 37. Right: normal factor graph $\mathbf{N}(\mathbf{y})$ that is discussed in Example 42.

- Let ψ be the surjective mapping

$$\psi : \mathcal{B} \rightarrow \mathcal{P}, \quad \beta \mapsto (\beta_{e,1})_{e \in \mathcal{E}_{\text{half}}}.$$

(Clearly, in general there are many $\beta \in \mathcal{B}$ that map to the same pseudo-codeword in \mathcal{P} .)

- Let Ψ_{BME} be the mapping

$$\Psi_{\text{BME}} : \mathcal{P} \rightarrow \mathcal{B}, \quad \omega \mapsto \arg \max_{\beta \in \mathcal{B}: \psi(\beta) = \omega} H_{\text{B}}(\beta), \quad (35)$$

where “BME” stands for “Bethe Max-Entropy.” This mapping gives for each $\omega \in \mathcal{P}$ the β among all the ψ -pre-images of ω that has the maximal Bethe entropy function value.

- The induced Bethe entropy function is defined to be

$$H_{\text{B}} : \mathcal{P} \rightarrow \mathbb{R}, \quad \omega \mapsto H_{\text{B}}(\Psi_{\text{BME}}(\omega)).$$

(Note that the argument of H_{B} determines if H_{B} denotes the Bethe entropy function or the induced Bethe entropy function.) \square

IX. GRAPH-COVER DECODING

As discussed in Section I, and shown in Figs. 1 and 2, graph-cover decoding is a theoretical tool to connect a variety of known decoders. In this section, we first review blockwise maximum a-posteriori decoding (BMAPD), which will set the stage for discussing blockwise graph-cover decoding (BGCD). Afterwards, we review symbolwise maximum a-posteriori decoding (SMAPD), upon which we introduce symbolwise graph-cover decoding (SGCD). These decoders are summarized in Tables I and II.

Note that blockwise graph-cover decoding was simply called graph-cover decoding in [5, Sec. 4] and that the exposition here is slightly more general than in [5] because we do not restrict ourselves to binary codes.

Definition 41 The setup in this section is as follows. (See also the upcoming Example 42.) We consider a discrete

TABLE I
EXPRESSIONS FOR THE PSEUDO-MARGINAL VECTORS THAT APPEAR IN A UNIFIED FORMULATION OF BMAPD, BGCD, SMAPD, AND SGCD
IN TERMS OF THE GLOBAL FUNCTION OF THE NFG $\mathbf{N} \triangleq \mathbf{N}(\mathbf{y})$ AND ITS FINITE GRAPH COVERS.
(FOR EVERY $M \in \mathbb{Z}_{>0}$, THE SCALAR $Z'_M(\mathbf{y}) \in \mathbb{R}_{>0}$ IS SOME SUITABLY DEFINED CONSTANT.)

	maximum a-posteriori decoding	graph-cover decoding
blockwise	$\hat{\beta}^{\text{BMAPD}}(\mathbf{y}) \triangleq \varphi_M \left(\arg \max_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M, \tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} \tilde{g}_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}) \right) \Big _{M=1}$	$\hat{\beta}^{\text{BGCD}}(\mathbf{y}) \triangleq \lim_{M \rightarrow \infty} \varphi_M \left(\arg \max_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M, \tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} \tilde{g}_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}) \right)$
symbolwise	$\hat{\beta}^{\text{SMAPD}}(\mathbf{y}) \triangleq \frac{1}{Z'_M(\mathbf{y})} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} \tilde{g}_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}) \cdot \varphi_M(\tilde{\mathbf{N}}, \tilde{\mathbf{c}}) \Big _{M=1}$	$\hat{\beta}^{\text{SGCD}}(\mathbf{y}) \triangleq \lim_{M \rightarrow \infty} \frac{1}{Z'_M(\mathbf{y})} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} \tilde{g}_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}) \cdot \varphi_M(\tilde{\mathbf{N}}, \tilde{\mathbf{c}})$

TABLE II
EXPRESSIONS FOR THE PSEUDO-MARGINAL VECTORS THAT APPEAR IN A UNIFIED FORMULATION OF BMAPD, BGCD, SMAPD, AND SGCD
IN TERMS OF GIBBS AND BETHE FREE ENERGY FUNCTIONS. HERE, IF $\beta \in \text{conv}(\mathcal{B}'_1)$, *i.e.*, β IS A GLOBALLY REALIZABLE PSEUDO-MARGINAL VECTOR
CORRESPONDING TO $\mathbf{p} \in \Pi_{\mathcal{C}}$, WE DEFINE $F_G(\beta) \triangleq F_G(\mathbf{p})$. (SEE ALSO REMARK 26.)

	maximum a-posteriori decoding	graph-cover decoding
blockwise	$\hat{\beta}^{\text{BMAPD}}(\mathbf{y}) = \arg \min_{\beta \in \text{conv}(\mathcal{B}'_1)} F_G(\beta) \Big _{T=0}$	$\hat{\beta}^{\text{BGCD}}(\mathbf{y}) = \arg \min_{\beta \in \mathcal{B}} F_B(\beta) \Big _{T=0}$
symbolwise	$\hat{\beta}^{\text{SMAPD}}(\mathbf{y}) = \arg \min_{\beta \in \text{conv}(\mathcal{B}'_1)} F_G(\beta) \Big _{T=1}$	$\hat{\beta}^{\text{SGCD}}(\mathbf{y}) = \arg \min_{\beta \in \mathcal{B}} F_B(\beta) \Big _{T=1}$

memoryless channel with an arbitrary input alphabet \mathcal{X} , an arbitrary output alphabet \mathcal{Y} , and arbitrary channel law $\{W(y|x)\}_{y \in \mathcal{Y}, x \in \mathcal{X}}$, *i.e.*, the probability of observing the symbol $y \in \mathcal{Y}$ at the channel output given that the symbol $x \in \mathcal{X}$ was sent is $W(y|x)$. Moreover, let \mathcal{C}_{ch} be a block code of length n and with alphabet \mathcal{X} that is used for data transmission over this discrete memoryless channel. We let $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$ be the random vectors corresponding to, respectively, the channel input and output symbols of n channel uses. We assume that a codeword $\mathbf{x} \in \mathcal{C}_{\text{ch}}$ is selected with probability $P_{\mathbf{X}}(\mathbf{x})$. (Of course, $P_{\mathbf{X}}(\mathbf{x}) = 0$ for $\mathbf{x} \notin \mathcal{C}_{\text{ch}}$.) The joint probability mass function of \mathbf{X} and \mathbf{Y} is then given by

$$P_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = P_{\mathbf{X}}(\mathbf{x}) \cdot P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = P_{\mathbf{X}}(\mathbf{x}) \cdot \prod_{i \in [n]} W(y_i|x_i).$$

For a given channel output vector $\mathbf{y} = (y_i)_{i \in [n]} \in \mathcal{Y}^n$, consider an NFG $\mathbf{N}(\mathbf{y}) \triangleq \mathbf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ with the following properties.

- For all $e \in \mathcal{E}_{\text{half}}$ we have $\mathcal{A}_e = \mathcal{X}$.
- We identify $\mathcal{E}_{\text{half}}$ with $[n]$.
- We identify $\{a_e\}_{e \in \mathcal{E}_{\text{half}}}$ with $\{x_i\}_{i \in [n]}$.
- In order to take the received vector \mathbf{y} into account, some function nodes are parameterized by y_i , $i \in [n]$.
- For every codeword $\mathbf{x} \in \mathcal{C}_{\text{ch}}$, there is exactly one valid configuration $\mathbf{c} \in \mathcal{C}$ such that the restriction of \mathbf{c} to $\mathcal{E}_{\text{half}}$ equals \mathbf{x} . This valid configuration will be denoted by $\mathbf{c}(\mathbf{x})$. In terms of Definition 36 this means that we impose $t_{\mathbf{N}} = 1$. (With the necessary care, the results of this section can be generalized to NFGs for which there exists a constant $t_{\mathbf{N}}$ with $t_{\mathbf{N}} > 1$.)
- There is some constant $\gamma \in \mathbb{R}_{>0}$ such that for every

codeword $\mathbf{x} \in \mathcal{C}_{\text{ch}}$, the global function value of the valid configuration $\mathbf{c}(\mathbf{x})$ is

$$g(\mathbf{c}(\mathbf{x})) = \gamma \cdot P(\mathbf{x}, \mathbf{y}). \quad (36)$$

□

Example 42 Consider again the code \mathcal{C}_{ch} from Example 37 (which was represented by the NFG in Fig. 11 (left)), along with some discrete memoryless channel with input alphabet $\mathcal{X} \triangleq \mathbb{F}_2$, output alphabet \mathcal{Y} , and channel law $\{W(y|x)\}_{y \in \mathcal{Y}, x \in \mathcal{X}}$. Let $\mathbf{y} \in \mathcal{Y}^n$ be a given channel output vector. It can be verified that Fig. 11 (right) shows a possible NFG $\mathbf{N}(\mathbf{y})$ that has the properties as specified in Definition 41.

□

Potential ties in upcoming “arg max” and “arg min” expressions are assumed to be resolved in a systematic or random manner.

A. Blockwise Maximum A-Posteriori Decoding

With the setup as in Definition 41, let $\hat{\mathbf{x}}^{\text{BMAPD}}(\mathbf{y}) \in \mathcal{X}^n$ be the decision vector obtained by BMAPD based on the received vector \mathbf{y} . (Recall that BMAPD is the decision rule that minimizes the block decision error probability $\Pr(\hat{\mathbf{x}}^{\text{BMAPD}}(\mathbf{Y}) \neq \mathbf{X})$.)

Definition 43 Given a channel output vector \mathbf{y} , BMAPD yields the decision rule

$$\begin{aligned} \hat{\mathbf{x}}^{\text{BMAPD}}(\mathbf{y}) &\triangleq \arg \max_{\mathbf{x}} P_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) \\ &= \arg \max_{\mathbf{x}} P_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}). \end{aligned} \quad (37)$$

On the side, we note that if all codewords are selected equally likely, i.e., $P(\mathbf{x}) = \frac{1}{|\mathcal{C}_{\text{ch}}|} \cdot [\mathbf{x} \in \mathcal{C}_{\text{ch}}]$, then this decision rule equals the blockwise maximum likelihood decoding rule.

Lemma 44 *Given a channel output vector \mathbf{y} , consider the NFG $\mathbf{N} \triangleq \mathbf{N}(\mathbf{y})$ from Definition 41. The vector $\hat{\mathbf{x}}^{\text{BMAPD}}(\mathbf{y})$ satisfies*

$$\hat{\mathbf{x}}^{\text{BMAPD}}(\mathbf{y}) = \arg \max_{\mathbf{c}_{\text{half}}} \max_{\mathbf{c}_{\text{full}}} g(\mathbf{c}_{\text{half}}, \mathbf{c}_{\text{full}})$$

and

$$\mathbf{c}(\hat{\mathbf{x}}^{\text{BMAPD}}(\mathbf{y})) = \arg \max_{\mathbf{c}} g(\mathbf{c}).$$

In terms of the pseudo-marginal vector $\hat{\beta}^{\text{BMAPD}}(\mathbf{y})$ that is defined in Table I, one can therefore write

$$[a_e = \hat{x}_e^{\text{BMAPD}}(\mathbf{y})] = \hat{\beta}_{e,a_e}^{\text{BMAPD}}(\mathbf{y}), \quad e \in \mathcal{E}_{\text{half}}, a_e \in \mathcal{A}_e,$$

and so

$$\hat{x}_e^{\text{BMAPD}}(\mathbf{y}) = \arg \max_{a_e} \hat{\beta}_{e,a_e}^{\text{BMAPD}}(\mathbf{y}), \quad e \in \mathcal{E}_{\text{half}}.$$

Proof: Follows from (36) and (37). Note that $\tilde{\mathcal{N}}_1 = \{\mathbf{N}(\mathbf{y})\}$. ■

As shown in the following lemma, the BMAPD rule can also be cast as a Gibbs free energy function minimization problem (with temperature $T = 0$).

Lemma 45 *Given a channel output vector \mathbf{y} , consider the NFG $\mathbf{N} \triangleq \mathbf{N}(\mathbf{y})$ from Definition 41 and define*

$$\hat{\mathbf{p}} \triangleq \arg \min_{\mathbf{p} \in \Pi_{\mathcal{C}}} F_{\mathbf{G}}(\mathbf{p}) \Big|_{T=0}.$$

The vector $\hat{\mathbf{p}}$ is such that

$$\hat{p}_{\mathbf{c}} = \begin{cases} 1 & \text{if } \mathbf{c} = \mathbf{c}(\hat{\mathbf{x}}^{\text{BMAPD}}(\mathbf{y})) \\ 0 & \text{otherwise} \end{cases}.$$

Proof: Follows from Definition 7, Eqs. (36) and (37), and the fact that $F_{\mathbf{G}}(\mathbf{p}) = U_{\mathbf{G}}(\mathbf{p})$ for temperature $T = 0$. ■

From Lemma 45, Remark 26, and the fact that $\tilde{\mathcal{N}}_1 = \{\mathbf{N}(\mathbf{y})\}$, it follows that $\hat{\beta}^{\text{BMAPD}}(\mathbf{y})$ can also be written as shown in Table II.

B. Blockwise Graph-Cover Decoding

We consider the setup as in Definition 41. Recall that BMAPD can be seen as a competition of all codewords to be the best explanation of the observed channel output vector. In this subsection we revisit blockwise graph-cover decoding (BGCD), which was originally introduced in [5, Section 4]. Actually, we will define this decoder slightly differently than in [5, Section 4]. Namely, whereas in [5, Section 4] all codewords in all finite covers of an NFG were competing to be the best explanation of a channel output vector, here we restrict the competition to all codewords in all M -covers of an NFG, and then we let M go to infinity.

□ **Definition 46** *Given a channel output vector \mathbf{y} , consider the NFG $\mathbf{N} \triangleq \mathbf{N}(\mathbf{y})$ from Definition 41. For any $M \in \mathbb{Z}_{>0}$, we define degree- M BGCD to be the decoding rule that gives back the pseudo-marginal vector*

$$\hat{\beta}^{\text{BGCD}(M)}(\mathbf{y}) \triangleq \varphi_M \left((\hat{\mathbf{N}}, \hat{\mathbf{c}})^{\text{BGCD}(M)}(\mathbf{y}) \right),$$

where

$$(\hat{\mathbf{N}}, \hat{\mathbf{c}})^{\text{BGCD}(M)}(\mathbf{y}) = \arg \max_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M, \tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} \tilde{g}_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}). \quad (38)$$

In the limit $M \rightarrow \infty$, we define BGCD to be the decoding rule that gives back the pseudo-marginal vector

$$\hat{\beta}^{\text{BGCD}}(\mathbf{y}) \triangleq \lim_{M \rightarrow \infty} \hat{\beta}^{\text{BGCD}(M)}(\mathbf{y}).$$

(This latter expression is also shown in Table I.) □

In the case $\mathcal{X} = \mathbb{F}_2$, one could have defined BGCD to give back the pseudo-codeword $\psi(\hat{\beta}^{\text{BGCD}}(\mathbf{y}))$ (with suitable generalizations for other alphabets \mathcal{X}), however, for simplicity of notation we will not pursue this option here.

Theorem 47 *Given a channel output vector \mathbf{y} , consider the NFG $\mathbf{N} \triangleq \mathbf{N}(\mathbf{y})$ from Definition 41. Then*

$$\hat{\beta}^{\text{BGCD}}(\mathbf{y}) = \arg \min_{\beta \in \mathcal{B}} F_{\mathbf{B}}(\beta) \Big|_{T=0}.$$

Proof: This follows from Theorems 25 and 31 and the fact that $F_{\mathbf{B}}(\beta) = U_{\mathbf{B}}(\beta)$ for temperature $T = 0$. ■

The decoder relationships that are highlighted in Fig. 1 are a consequence of the following observations.

- Finding the minimum of the Bethe free energy function at temperature $T = 0$ is equivalent to linear programming decoding.
- As shown in Theorem 47, blockwise graph-cover decoding is equivalent to finding the minimum of the Bethe free energy function at temperature $T = 0$.
- As discussed in [4], [5] and in Section I, a locally operating algorithm like the max-product (min-sum) algorithm “cannot distinguish” if it is operating on an NFG \mathbf{N} or, implicitly, on any of its covers. (In particular, note that the fact that any finite graph cover $\tilde{\mathbf{N}}$ of \mathbf{N} looks locally the same as \mathbf{N} implies that the collection of computation trees of \mathbf{N} equals the collection of computation trees of $\tilde{\mathbf{N}}$.) With this, BGCD can be considered to be a “model” for the behavior of max-product (min-sum) algorithm decoding. Note that the connection between BGCD and max-product (min-sum) algorithm decoding is in general only an approximate one. However, in all cases where analytical tools are known that exactly characterize the behavior of the max-product algorithm decoder, the connection between the BGCD and the max-product (min-sum) algorithm decoder is exact.

Note that if the NFG \mathbf{N} does not contain cycles then max-product algorithm decoding and linear programming decoding yield the same decision as BMAPD [50]. This is reflected in the equivalence of $F_{\mathbf{G}}$ and $F_{\mathbf{B}}$ for cycle-free NFGs, once the domains of these two functions have been suitably identified.

C. Symbolwise Maximum A-Posteriori Decoding

With the setup as in Definition 41, let $\hat{\mathbf{x}}^{\text{SMAPD}}(\mathbf{y}) \in \mathcal{X}^n$ be the decision vector obtained by SMAPD based on the received vector \mathbf{y} . Recall that SMAPD is the decision rule that minimizes the symbol decision error probability $\Pr(\hat{x}_i^{\text{SMAPD}}(\mathbf{Y}) \neq X_i)$ for every $i \in [n]$ (or, depending on the definition, for every i that corresponds to an information symbol of \mathcal{C}_{ch}).

Definition 48 Given a channel output vector \mathbf{y} , SMAPD yields the vector $\hat{\mathbf{x}}^{\text{SMAPD}}(\mathbf{y})$ with components

$$\begin{aligned} \hat{x}_i^{\text{SMAPD}}(\mathbf{y}) &\triangleq \arg \max_{x_i} P_{X_i|\mathbf{Y}}(x_i|\mathbf{y}) \\ &= \arg \max_{x_i} P_{X_i, \mathbf{Y}}(x_i, \mathbf{y}), \quad i \in [n], \end{aligned}$$

where

$$P_{X_i, \mathbf{Y}}(x_i, \mathbf{y}) = \sum_{\mathbf{x}': x'_i = x_i} P_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}', \mathbf{y}). \quad (39)$$

□

Note that $\hat{\mathbf{x}}^{\text{SMAPD}}(\mathbf{y})$, in contrast to $\hat{\mathbf{x}}^{\text{BMAPD}}(\mathbf{y})$, is not always a codeword.

Lemma 49 Given a channel output vector \mathbf{y} , consider the NFG $\mathbf{N} \triangleq \mathbf{N}(\mathbf{y})$ from Definition 41. The SMAPD vector $\hat{\mathbf{x}}^{\text{SMAPD}}(\mathbf{y})$ satisfies

$$\hat{x}_e^{\text{SMAPD}}(\mathbf{y}) = \arg \max_{a_e} \eta_e(a_e), \quad e \in \mathcal{E}_{\text{half}}. \quad (40)$$

with

$$\eta_e(a_e) \triangleq \frac{1}{Z_G} \cdot \sum_{\mathbf{a}': a'_e = a_e} g(\mathbf{a}'), \quad e \in \mathcal{E}_{\text{half}}, a_e \in \mathcal{A}_e. \quad (41)$$

In terms of the pseudo-marginal vector $\hat{\beta}^{\text{SMAPD}}(\mathbf{y})$ that is defined in Table I, one can therefore write

$$\eta_e(a_e) = \hat{\beta}_{e, a_e}^{\text{SMAPD}}(\mathbf{y}), \quad e \in \mathcal{E}_{\text{half}}, a_e \in \mathcal{A}_e,$$

and

$$\hat{x}_e^{\text{SMAPD}}(\mathbf{y}) = \arg \max_{a_e} \hat{\beta}_{e, a_e}^{\text{SMAPD}}(\mathbf{y}), \quad e \in \mathcal{E}_{\text{half}}.$$

Proof: Follows from (36) and (39). Note that $\tilde{\mathcal{N}}_1 = \{\mathbf{N}(\mathbf{y})\}$. ■

From Lemma 49 it is clear that the main computational step towards obtaining the SMAPD vector is to compute the marginals $\{\eta_e(a_e)\}_{e \in \mathcal{E}_{\text{half}}, a_e \in \mathcal{A}_e}$. Note that the scaling constant in (41) was chosen such that $\sum_{a_e} \eta_e(a_e) = 1$ for every $e \in \mathcal{E}_{\text{half}}$. Of course, any other positive scaling factor works equally well as it entails the same decision in (40).

As shown in the following lemma, the SMAPD rule can also be cast as a Gibbs free energy function minimization problem (with temperature $T = 1$).

Lemma 50 Given a channel output vector \mathbf{y} , consider the NFG $\mathbf{N} \triangleq \mathbf{N}(\mathbf{y})$ from Definition 41 and define

$$\hat{\mathbf{p}} \triangleq \arg \min_{\mathbf{p} \in \Pi_{\mathcal{C}}} F_G(\mathbf{p}) \Big|_{T=1}.$$

Using the notation from Lemma 49, we have

$$\eta_e(a_e) = \sum_{\mathbf{c}: c_e = a_e} \hat{\mathbf{p}}_{\mathbf{c}}, \quad e \in \mathcal{E}_{\text{half}}, a_e \in \mathcal{A}_e.$$

Proof: Follows from Definition 7 and Lemmas 8 and 49. ■

From Lemmas 49 and 50, Remark 26, and the fact that $\tilde{\mathcal{N}}_1 = \{\mathbf{N}(\mathbf{y})\}$, it follows that $\hat{\beta}^{\text{BMAPD}}(\mathbf{y})$ can also be written as shown in Table II.

D. Symbolwise Graph-Cover Decoding

We consider the setup as in Definition 41. Recall that SMAPD is based on computing suitable marginals of the global function represented by the NFG $\mathbf{N}(\mathbf{y})$. In this subsection we define symbolwise graph-cover decoding (SGCD), which was outlined in [64], [65]. Similar to the transition from BMAPD to BGCD, where the competition is extended from all codewords of $\mathbf{N}(\mathbf{y})$ to all codewords in all M -covers of $\mathbf{N}(\mathbf{y})$, when going from SMAPD to SGCD we replace the marginals of the global function of $\mathbf{N}(\mathbf{y})$ by a suitable combination of marginals of the global functions of all M -covers of $\mathbf{N}(\mathbf{y})$.

Definition 51 Given a channel output vector \mathbf{y} , consider the NFG $\mathbf{N} \triangleq \mathbf{N}(\mathbf{y})$ from Definition 41. For any $M \in \mathbb{Z}_{>0}$, we define degree- M SGCD to yield the pseudo-marginal vector $\hat{\beta}^{\text{SGCD}(M)}(\mathbf{y})$ with components

$$\begin{aligned} \hat{\beta}_{e, a_e}^{\text{SGCD}(M)}(\mathbf{y}) &\triangleq \eta_{e, M}(a_e), \quad e \in \mathcal{E}, a_e \in \mathcal{A}_e, \\ \hat{\beta}_{f, \mathbf{a}_f}^{\text{SGCD}(M)}(\mathbf{y}) &\triangleq \eta_{f, M}(\mathbf{a}_f), \quad f \in \mathcal{F}, \mathbf{a}_f \in \mathcal{A}_f. \end{aligned}$$

For every $e \in \mathcal{E}$, the “marginal function” $\eta_{e, M}$ is defined by

$$\begin{aligned} \eta_{e, M}(a_e) &\triangleq \frac{1}{M} \cdot \sum_{m \in [M]} \eta_{e, m, M}(a_e), \\ e &\in \mathcal{E}, a_e \in \mathcal{A}_e, \end{aligned}$$

$$\begin{aligned} \eta_{e, m, M}(a_e) &\triangleq \frac{1}{Z'_M(\mathbf{N})} \cdot \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} Z_G(\tilde{\mathbf{N}}) \cdot \eta_{e, m, \tilde{\mathbf{N}}}(a_e), \\ e &\in \mathcal{E}, m \in [M], a_e \in \mathcal{A}_e, \end{aligned}$$

$$\begin{aligned} \eta_{e, m, \tilde{\mathbf{N}}}(a_e) &\triangleq \frac{1}{Z_G(\tilde{\mathbf{N}})} \cdot \sum_{\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}}): \tilde{c}_{e, m} = a_e} g_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}), \\ e &\in \mathcal{E}, m \in [M], a_e \in \mathcal{A}_e, \end{aligned}$$

where

$$Z'_M(\mathbf{N}) \triangleq \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} Z_G(\tilde{\mathbf{N}}) = |\tilde{\mathcal{N}}_M| \cdot (Z_{\mathbf{B}, M}(\mathbf{N}))^M. \quad (42)$$

(For a motivation of these expressions, see the paragraph after this definition.) For every $f \in \mathcal{F}$, the “marginal function” $\eta_{f, M}$ is defined analogously. Moreover, taking the limit $M \rightarrow \infty$, we define SGCD to be the decoder that gives back the pseudo-marginal vector

$$\hat{\beta}^{\text{SGCD}}(\mathbf{y}) \triangleq \lim_{M \rightarrow \infty} \hat{\beta}^{\text{SGCD}(M)}(\mathbf{y}).$$

□

Here are the motivations for specifying the “marginal functions” as we did in Definition 51.

- Fix an arbitrary M -cover \tilde{N} of N . If we were to find the SMAPD estimate of the code defined by \tilde{N} then, following Lemma 49, we would have to compute the marginals

$$\{\eta_{e,m,\tilde{N}}(\tilde{a}_{e,m})\}_{e \in \mathcal{E}_{\text{half}}, m \in [M], \tilde{a}_{e,m} \in \mathcal{A}_e}.$$

- However, given the fact that no M -cover is more special than any other M -cover, we take the average of these marginals over all M -covers and obtain the marginals

$$\{\eta_{e,m,M}(a_e)\}_{e \in \mathcal{E}_{\text{half}}, m \in [M], a_e \in \mathcal{A}_e}.$$

Actually, we take a weighted average where the weighting factor for \tilde{N} is chosen to be its Gibbs partition function $Z_G(\tilde{N})$.

- From the symmetries of the setup it is clear that for every $e \in \mathcal{E}$ and every $a_e \in \mathcal{A}_e$ the quantity $\eta_{e,m,M}(a_e)$ is independent of $m \in [M]$. Therefore, the definition of the marginals

$$\{\eta_{e,M}(a_e)\}_{e \in \mathcal{E}_{\text{half}}, a_e \in \mathcal{A}_e}$$

is somewhat trivial, but notationally and analytically useful.

- The scaling factors were chosen such that the marginals sum to 1, when summed over their corresponding alphabets. (Note that the reformulation of $Z'_M(N)$ on the right-hand side of (42) follows from Definition 32.)

Theorem 52 *Given a channel output vector \mathbf{y} , consider the NFG $N(\mathbf{y})$ from Definition 41 and define*

$$\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} F_B(\beta) \Big|_{T=1}.$$

Using the notation from Definition 51, we have

$$\begin{aligned} \lim_{M \rightarrow \infty} \eta_{e,M}(a_e) &= \hat{\beta}_{e,a_e}, \quad e \in \mathcal{E}, \quad a_e \in \mathcal{A}_e, \\ \lim_{M \rightarrow \infty} \eta_{f,M}(\mathbf{a}_f) &= \hat{\beta}_{f,\mathbf{a}_f}, \quad f \in \mathcal{F}, \quad \mathbf{a}_f \in \mathcal{A}_f. \end{aligned}$$

Proof: See Appendix C. ■

In the rather special case where F_B has multiple global minima (necessarily of equal value), Theorem 52 has to be stated somewhat more carefully. Namely, the pseudo-marginal vector $\hat{\beta}$ has to be replaced by a suitable pseudo-marginal vector in the convex hull of all pseudo-marginal vectors that minimize F_B .

The decoder relationships that are highlighted in Fig. 2 are a consequence of the following observations.

- As shown in Theorem 52, symbolwise graph-cover decoding is equivalent to finding the minimum of the Bethe free energy function at temperature $T = 1$.
- As discussed in [4], [5] and in Section I, a locally operating algorithm like the sum-product algorithm “cannot distinguish” if it is operating on an NFG N or, implicitly, on any of its covers. (Again, note that the fact that any finite graph cover \tilde{N} of N looks locally the same as N implies that the collection of computation trees of \tilde{N})

Therefore, SGCD can be considered to be a “model” for the behavior of sum-product algorithm decoding. Note that the connection between SGCD and SPA decoding is in general only an approximate one. However, in many cases where analytical tools are known that exactly characterize the behavior of the SPA decoder, the connection between SGCD and SPA decoding is exact.

Note that if the NFG $N(\mathbf{y})$ does not contain cycles then sum-product algorithm decoding yields the same (pseudo-)marginal vector as SMAPD. This is reflected in the equivalence of F_G and F_B for cycle-free NFGs, once the domains of these two functions have been suitably identified.

For an NFG *without* cycles, the meaning of the pseudo-marginal functions that are computed by the sum-product algorithm is clear (*cf.* the discussion at the beginning of Section I), but for an NFG *with* cycles, the meaning of these pseudo-marginal functions is a priori less clear. However, combining Theorems 30 and 52 with the theorem by Yedidia, Freeman, and Weiss [3] on the characterization of fixed points of the sum-product algorithm, one obtains the following statement. Namely, a fixed point of the sum-product algorithm corresponds to a certain pseudo-marginal vector of the factor graph under consideration: it is, after taking a biasing channel-output-dependent term properly into account, the pseudo-marginal vector that has (locally) an extremal number of pre-images in all M -covers, when M goes to infinity.⁵

X. THE INFLUENCE OF THE MINIMUM HAMMING DISTANCE OF A CODE UPON THE BETHE ENTROPY FUNCTION OF ITS NFG

It can easily be verified that the Gibbs entropy function is a concave function of its arguments. However, the Bethe entropy function is in general *not* a concave function of its arguments. This has important consequences when trying to minimize the Bethe free energy function because the curvature of this function is determined by the curvature of the Bethe entropy function.

In this section we show that choosing a code from an ensemble of regular LDPC codes with minimum Hamming distance growing (with high probability) linearly with the block length comes at the price of having to deal with an NFG whose Bethe entropy function is concave and convex and whose Bethe free energy function is, therefore, convex and concave. (By a multidimensional function being “concave and convex” we mean that there are points and directions where the function is locally concave, and points and directions where the function is locally convex.) Moreover, we show that the choice of a code from such an ensemble has implications for the accuracy of the pseudo-marginals that are computed by the sum-product algorithm.

⁵In this statement we included the word “locally” because the sum-product algorithm can get stuck at a local extremum of the Bethe free energy function. Note that here the use of the word “local” is different than the use of the word “local” when comparing the global perspective of maximum a-posteriori decoding with the local perspective of message-passing iterative decoding. However, ultimately, this “local” is also a consequence of the suboptimal behavior of message-passing iterative decoding stemming from its local perspective.

We conjecture that the above results are also valid for ensembles of *irregular* LDPC codes whose minimum Hamming distance grows (with high probability) linearly with the block length, however, we prove the above statements only for the case of regular LDPC codes.

This section is structured as follows. In Section X-A we make a simple observation about the induced Bethe entropy function for regular LDPC codes. Afterwards, in Section X-B we discuss how this observation implies the above-mentioned results.

A. An Observation about the Induced Bethe Entropy Function

In this subsection we consider a setup where the expression for the induced Bethe entropy (see Definition 40) can be simplified significantly. Afterwards we will recognize that the obtained expression appears also in some other context. This will then lead to the promised conclusions, which are discussed in the next subsection.

Recall the definition of a (d_L, d_R) -regular LDPC code from Definition 38. Note that the rate of such a code is lower bounded by $1 - d_L/d_R$.

Lemma 53 *Consider a (d_L, d_R) -regular length- n LDPC code over \mathbb{F}_2 described by some parity-check matrix \mathbf{H} , and let $N(\mathbf{H})$ be the NFG associated with \mathbf{H} as in Definition 38. Then the induced Bethe entropy function along the straight line*

$$\omega(s) \triangleq \omega(s) \cdot (1, \dots, 1), \quad s \in \mathbb{R},$$

evaluates to

$$H_B(\omega(s)) = n \cdot h_{d_L, d_R}(s), \quad s \in \mathbb{R}.$$

Here we have used the functions

$$\begin{aligned} h_{d_L, d_R} : \mathbb{R} &\rightarrow \mathbb{R} \\ s &\mapsto -(d_L - 1) \cdot h(\omega(s)) - d_L \cdot s \cdot \omega(s) + \frac{d_L}{d_R} \cdot \theta(s), \\ \omega : \mathbb{R} &\rightarrow \mathbb{R}, \quad s \mapsto \frac{1}{d_R} \cdot \frac{d}{ds} \theta(s), \\ \theta : \mathbb{R} &\rightarrow \mathbb{R}, \quad s \mapsto \log \left(\sum_{\substack{w=0 \\ w \text{ even}}}^{d_R} \binom{d_R}{w} \exp(s \cdot w) \right), \\ h : [0, 1] &\rightarrow \mathbb{R}, \quad \xi \mapsto -\xi \log(\xi) - (1 - \xi) \log(1 - \xi). \end{aligned}$$

Proof: See Appendix D. ■

Example 54 *For $(d_L, d_R) = (2, 4)$ and $(d_L, d_R) = (3, 6)$ the graph of $s \mapsto (\omega(s), h_{d_L, d_R}(s))$ is visualized in Figs. 12 and 13, respectively. We make the following observations with respect to the shapes of these curves and the values that $h_{d_L, d_R}(s)$ takes on.*

- In the case $(d_L, d_R) = (2, 4)$, it can be verified that the graph $s \mapsto (\omega(s), h_{d_L, d_R}(s))$ is concave and that $h_{d_L, d_R}(s)$ is always non-negative.
- In the case $(d_L, d_R) = (3, 6)$, it can be verified that the graph $s \mapsto (\omega(s), h_{d_L, d_R}(s))$ is concave and convex.

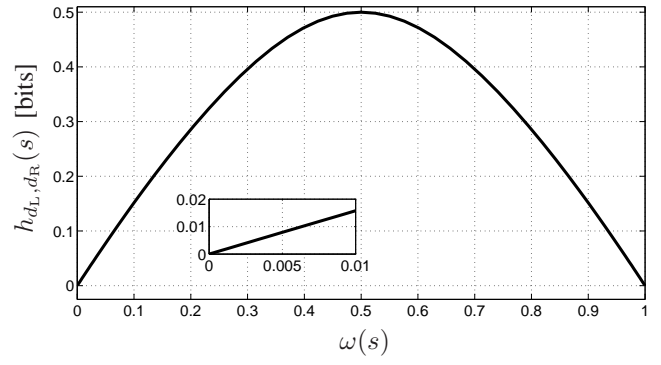


Fig. 12. The graph of $s \mapsto (\omega(s), h_{d_L, d_R}(s))$ for $(d_L, d_R) = (2, 4)$. The inset zooms into the area near the origin.

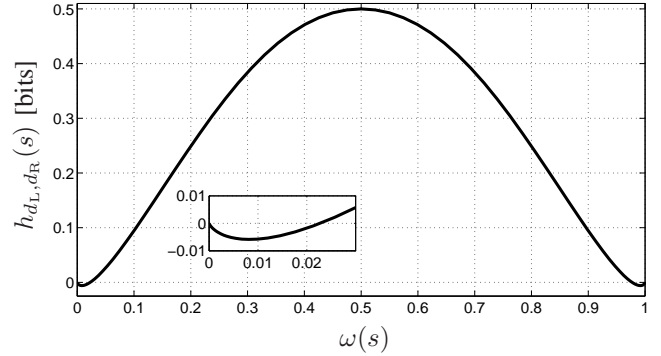


Fig. 13. The graph of $s \mapsto (\omega(s), h_{d_L, d_R}(s))$ for $(d_L, d_R) = (3, 6)$. The inset zooms into the area near the origin.

Moreover, for small $\omega(s) > 0$ the value of $h_{d_L, d_R}(s)$ is negative. (This behavior is actually typical for the behavior of $s \mapsto (\omega(s), h_{d_L, d_R}(s))$ for any (d_L, d_R) -regular LDPC code with $3 \leq d_L < d_R$.) □

It is worth emphasizing that the expression for $H_B(\omega(s))$ in Lemma 53 holds for any (d_L, d_R) -regular LDPC code over \mathbb{F}_2 of length n , i.e., it is neither an ensemble average result, nor an asymptotic (in n) result.

Remark 55 *Interestingly enough, the functions ω and h_{d_L, d_R} from Lemma 53 appear also when studying the ensemble of (d_L, d_R) -regular LDPC codes with block length going to infinity. Namely, the asymptotic growth rate of the average number of codewords of relative Hamming weight $\omega(s)$ is given by $h_{d_L, d_R}(s)$, where the average is taken over Gallager's ensemble of (d_L, d_R) -regular LDPC codes [66, Section 2]. (The same asymptotic growth rate is also obtained for the ensemble of all (d_L, d_R) -regular LDPC codes as defined by Richardson and Urbanke [70], see also [67], [71], [72].) □*

Using the interpretation of the Bethe entropy function that was given in Section VII, this equivalence is not totally surprising considering the following facts. (Here, $N(\mathbf{H})$ refers to the NFG in Lemma 53.)

- Because $N(\mathbf{H})$ represents a (d_L, d_R) -regular LDPC code, any finite graph cover of $N(\mathbf{H})$ also represents a (d_L, d_R) -regular LDPC code.

- A “typical” codeword of relative Hamming weight $\omega(s)$ in a finite graph cover of $N(\mathbf{H})$ maps down to a pseudo-codeword that is very close to $\omega(s) \cdot (1, \dots, 1)$.
- The Bethe entropy function value of some pseudo-marginal vector in the local marginal polytope of $N(\mathbf{H})$ “counts” the number of valid configurations in finite graph covers of $N(\mathbf{H})$ that map down to that pseudo-marginal vector. (See Section VII for a more precise statement.)

We leave it as an open problem to find a suitable generalization of Lemma 53 to irregular LDPC codes and ensembles of irregular LDPC codes.

B. Implications of the Above Observation

In this subsection we explore some of the implications of the observation that the graph $s \mapsto (\omega(s), h_{d_L, d_R}(s))$ appears in two different setups, namely in the setup of Lemma 53 and in the setup of Remark 55. For this discussion, recall that a function with a multi-dimensional domain is called concave if at every point of its domain the function is concave in every direction.

Let us first consider Gallager’s ensemble of (d_L, d_R) -regular LDPC codes where $3 \leq d_L < d_R$. It was already observed by Gallager [66] that codes from this ensemble have a minimum Hamming distance that grows (with high probability) linearly with the block length. A necessary condition for this to happen is that the function $h_{d_L, d_R}(s)$ is *negative* for small $\omega(s) > 0$ (see Fig. 13 for the case $(d_L, d_R) = (3, 6)$). Because $h_{d_L, d_R}(s) = 0$ for $\omega(s) = 0$ and because $h_{d_L, d_R}(s) > 0$ for sufficiently large $0 < \omega(s) < 1$, the function $h_{d_L, d_R}(s)$ must be a *convex* function of $\omega(s)$ for small $\omega(s)$. Combining this observation with Lemma 53 and Remark 55 yields the conclusion that the induced Bethe entropy function of an NFG of a code from this ensemble is concave and convex. (See the beginning of this section for our definition of “concave and convex.”) A slightly more involved analysis then also yields the conclusion that the Bethe entropy function of an NFG of a code from this ensemble is concave and convex.

Still talking about Gallager’s ensemble of (d_L, d_R) -regular LDPC codes where $3 \leq d_L < d_R$, these observations have also consequences for the computation of pseudo-marginal vectors with the help of fixed points of the sum-product algorithm. Recall that the theorem by Yedidia, Freeman, and Weiss [3] showed that fixed points of the sum-product algorithm correspond to stationary points of the Bethe free energy function. Now, the fact that the Bethe entropy function is not concave everywhere implies that the Bethe free energy function is not convex everywhere, in particular it is not convex in the vicinity of pseudo-marginal vectors that correspond to codewords. For continuing our argument, assume that the received vector is such that the true marginal vector is close to the marginal vector corresponding to some codeword. In order for the sum-product algorithm to be able to somewhat closely reproduce this true marginal vector, the sum-product algorithm would have to have a stable fixed point with a pseudo-marginal vector somewhat close to this marginal vector, *i.e.*, the Bethe free energy function would need to have a local minimum at a pseudo-marginal vector somewhat close to this

marginal vector. However, the above non-convexity results of the Bethe free energy function show that this is not possible for every true marginal vector.⁶ In conclusion, the accuracy of the sum-product-algorithm-based estimation of marginal vectors of NFGs of regular LDPC codes from ensembles with linearly growing minimum Hamming distance has its limitations. However, if only 0-vs.-1 decisions are important (as it very often is the case in channel coding theory) then these limitations are usually not that severe.

For completeness, let us also briefly discuss Gallager’s ensemble of (d_L, d_R) -regular LDPC codes where $2 = d_L < d_R$. As pointed out by Gallager [66], codes from this ensemble have a minimum Hamming distance that grows at most logarithmically with the block length. This is also reflected by the fact that $h_{d_L, d_R}(s)$ is *positive* for small $\omega(s) > 0$ (see Fig. 13 for the case $(d_L, d_R) = (2, 4)$). (This statement is not strong enough to prove concavity of $h_{d_L, d_R}(s)$ in $\omega(s)$. For establishing this, a detailed analysis of $h_{d_L, d_R}(s)$ as a function of $\omega(s)$ is necessary.)

XI. CONCLUSIONS

We have shown that it is possible to give a *combinatorial* characterization of the Bethe entropy function and the Bethe partition function, two functions that were originally defined *analytically*. The key was to study finite graph covers of the NFG under consideration, in particular to count valid configurations in these finite graph covers. Moreover, we have introduced a theoretical tool called symbolwise graph-cover decoding that helps to better understand the meaning of the pseudo-marginal vector at fixed points of the sum-product algorithm. For all these results, the main mathematical tool that we used was the method of types.

We finish with a few remarks.

- It is clear that all the results that were stated in this paper for temperature $T = 1$ can be suitably generalized to any temperature $T \in \mathbb{R}_{>0}$.
- The fractional Bethe approximation (see, *e.g.*, [73]) and the Kikuchi approximation (see, *e.g.*, [3], [74]) are usually better approximations than the Bethe approximation. Generalizing the results of the present paper, we have outlined a combinatorial characterization of the entropy function of these approximations in [75].
- Although the main application of symbolwise graph-cover decoding is in obtaining a better understanding of fixed points of the sum-product algorithm, one wonders if also the transient and the periodic behavior of the sum-product algorithm can be characterized in terms of graph covers (or variations thereof). Some initial results in that direction were sketched in [76].
- It might be interesting to study the influence of redundant parity-checks of LDPC codes upon the Bethe entropy function.

⁶In fact, the operation of the sum-product algorithm on the NFG of a code from this ensemble behaves such that once it has “locked into” some codeword, the pseudo-marginal vector produced by the sum-product algorithm is getting closer and closer to the marginal vector corresponding to that codeword.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge discussions with Ralf Koetter and Andi Loeliger on factor-graph-related topics. Moreover, we greatly appreciate the reviewers' and the associate editor's constructive comments that lead to an improved presentation of the results.

APPENDIX A
PROOF OF LEMMA 29

Similar to the proof of Lemma 20, a possibility to draw an M -cover of \mathbf{N} is to first draw M copies of every function node of \mathbf{N} , then to draw edges that suitably connect these function nodes, and finally, where required, to attach half-edges to the function nodes.

In this proof, we will use this drawing procedure to guide the counting process. Namely, we start by drawing M copies for each function node, along with the sockets where later on the edges will be attached to. Let us count in how many ways we can specify $\{\tilde{\alpha}_{f,m}\}_{f \in \mathcal{F}, m \in [M]}$ that are consistent with β . (We will call these locally valid configurations.) It can easily be seen that there are $\prod_{f \in \mathcal{F}} \binom{M}{M \cdot \beta_f}$ ways to do this.

Fix such a locally valid configuration that is consistent with β . Let us count how many graph covers $\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M$ specify an edge connection such that this locally valid configuration induces a valid configuration in $\tilde{\mathbf{N}}$. By this we mean that there is a unique valid configuration $\tilde{c} \in \mathcal{C}(\tilde{\mathbf{N}})$ such that the following holds.

- For every full-edge $(e, m) \in \mathcal{E}_{\text{full}} \times [M]$ we have

$$\tilde{\alpha}_{f',m',e,m} = \tilde{c}_{e,m} = \tilde{\alpha}_{f'',m'',e,m}.$$

(Here we assumed that the full-edge (e, m) connects the two function nodes (f', m') and (f'', m'') .)

- For every half-edge $(e, m) \in \mathcal{E}_{\text{half}} \times [M]$ we have

$$\tilde{c}_{e,m} = \tilde{\alpha}_{f,m,e,m}.$$

(Here we assumed that the half-edge (e, m) is connected to the function node (f, m) .)

There are precisely $\prod_{e \in \mathcal{E}_{\text{full}}} \prod_{a_e} (M \beta_{e,a_e})!$ such M -covers. This can be seen as follows. Namely, consider some full-edge $e \in \mathcal{E}_{\text{full}}$ that connects the two function nodes f' and f'' and fix some $a_e \in \mathcal{A}_e$. It follows from the edge consistency constraints of the local marginal polytope that β is such that

$$\sum_{\alpha'_{f'} \in \mathcal{A}_{f'} : \alpha'_{f',e} = a_e} \beta_{f',\alpha'_{f'}} = \beta_{e,a_e} = \sum_{\alpha''_{f''} \in \mathcal{A}_{f''} : \alpha''_{f'',e} = a_e} \beta_{f'',\alpha''_{f''}}.$$

Therefore, the number of e -sockets among the M copies of f' that take on the value a_e is $M \cdot \beta_{e,a_e}$, and the number of e -sockets among the M copies of f'' that take on the value a_e is also $M \cdot \beta_{e,a_e}$. These sockets can be connected by $M \cdot \beta_{e,a_e}$ edges in exactly $(M \beta_{e,a_e})!$ ways.

Note that the number $\prod_{e \in \mathcal{E}_{\text{full}}} \prod_{a_e} (M \beta_{e,a_e})!$ is independent of the chosen locally valid configuration, and so, among all M -covers, the total number of valid configurations that map down to β equals $\prod_{f \in \mathcal{F}} \binom{M}{M \cdot \beta_f} \cdot \prod_{e \in \mathcal{E}_{\text{full}}} \prod_{a_e} (M \beta_{e,a_e})!$. The lemma statement is then obtained by dividing this number by $|\tilde{\mathcal{N}}_M|$, by using the result in (15), and by applying the abbreviations that are defined in (22)–(23).

APPENDIX B
PROOF OF THEOREM 33

We start by reformulating the M th power of $Z_{\mathbf{B},M}(\mathbf{N})$. Namely, we have

$$\begin{aligned} (Z_{\mathbf{B},M}(\mathbf{N}))^M &\stackrel{(a)}{=} \left\langle Z_{\mathbf{G}}(\tilde{\mathbf{N}}) \right\rangle_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \\ &\stackrel{(b)}{=} \frac{1}{|\tilde{\mathcal{N}}_M|} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{c} \in \mathcal{C}(\tilde{\mathbf{N}})} g_{\tilde{\mathbf{N}}}(\tilde{c})^{1/T} \\ &\stackrel{(c)}{=} \sum_{\beta \in \mathcal{B}'_M} \frac{1}{|\tilde{\mathcal{N}}_M|} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{c} \in \mathcal{C}(\tilde{\mathbf{N}})} [\varphi_M(\tilde{\mathbf{N}}, \tilde{c}) = \beta] \cdot g_{\tilde{\mathbf{N}}}(\tilde{c})^{1/T} \\ &\stackrel{(d)}{=} \sum_{\beta \in \mathcal{B}'_M} \exp(-(M/T) \cdot U_{\mathbf{B}}(\beta)) \\ &\quad \cdot \frac{1}{|\tilde{\mathcal{N}}_M|} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{c} \in \mathcal{C}(\tilde{\mathbf{N}})} [\varphi_M(\tilde{\mathbf{N}}, \tilde{c}) = \beta] \\ &\stackrel{(e)}{=} \sum_{\beta \in \mathcal{B}'_M} \exp(-(M/T) \cdot U_{\mathbf{B}}(\beta)) \cdot \bar{C}_M(\beta) \\ &\stackrel{(f)}{=} \sum_{\beta \in \mathcal{B}'_M} \exp(-(M/T) \cdot F_{\mathbf{B}}(\beta) + o(M)), \end{aligned} \quad (43)$$

where at step (a) we have used Definition 32, where at step (b) we have used (4) and Definition 27, where at step (c) we have used Definitions 22 and 24, where at step (d) we have used Theorem 31, where at step (e) we have used Eq. (21) and Definitions 27 and 28, and where at step (f) we have used (24). Consequently we obtain

$$\begin{aligned} \limsup_{M \rightarrow \infty} Z_{\mathbf{B},M}(\mathbf{N}) &\stackrel{(a)}{=} \limsup_{M \rightarrow \infty} \sqrt[M]{\sum_{\beta \in \mathcal{B}'_M} \exp(-(M/T) \cdot F_{\mathbf{B}}(\beta) + o(M))} \\ &\stackrel{(b)}{=} \limsup_{M \rightarrow \infty} \sqrt[M]{\max_{\beta \in \mathcal{B}'_M} \exp(-(M/T) \cdot F_{\mathbf{B}}(\beta) + o(M))} \\ &\stackrel{(c)}{=} \limsup_{M \rightarrow \infty} \max_{\beta \in \mathcal{B}'_M} \exp(-(1/T) \cdot F_{\mathbf{B}}(\beta) + o(1)) \\ &\stackrel{(d)}{=} \sup_{\beta \in \mathcal{B}'} \exp(-(1/T) \cdot F_{\mathbf{B}}(\beta)) \\ &= \exp\left(-\frac{1}{T} \cdot \inf_{\beta \in \mathcal{B}'} F_{\mathbf{B}}(\beta)\right) \\ &\stackrel{(e)}{=} \exp\left(-\frac{1}{T} \cdot \min_{\beta \in \mathcal{B}} F_{\mathbf{B}}(\beta)\right) \\ &\stackrel{(f)}{=} Z_{\mathbf{B}}(\mathbf{N}), \end{aligned}$$

where at step (a) we have used (43), where in step (b) the replacement of the sum by the maximum operator is justified by the fact that the size of the set \mathcal{B}'_M grows polynomially in M (see Remark 26), where step (c) follows from taking the maximization operator out of the M th root, where in step (d) we have used the definition of \mathcal{B}' , the set of lift-realizable pseudo-marginal vectors, where at step (e) we have used the fact that the closure of \mathcal{B}' equals \mathcal{B} , which is a consequence of Theorem 25, and the fact that \mathcal{B} is compact so that the infimum operator can be replaced by a minimization operator,

and where at step (f) we have used Definition 15. This is the result that was promised in the theorem statement.

APPENDIX C PROOF OF THEOREM 52

From the derivations in this appendix it will be apparent that there are close connections to the proof of Theorem 33 in Appendix B.

For any $e \in \mathcal{E}$, any $M \in \mathbb{Z}_{>0}$, and any $a_e \in \mathcal{A}_e$ we have

$$\begin{aligned}
& \eta_{e,M}(a_e) \\
& \stackrel{(a)}{=} \frac{1}{M} \sum_{m \in [M]} \eta_{e,m,M}(a_e) \\
& \stackrel{(b)}{=} \frac{1}{M} \sum_{m \in [M]} \frac{1}{Z'_M(\mathbf{N})} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} Z_G(\tilde{\mathbf{N}}) \cdot \eta_{e,m,\tilde{\mathbf{N}}}(a_e) \\
& \stackrel{(c)}{=} \frac{1}{M} \sum_{m \in [M]} \frac{1}{Z'_M(\mathbf{N})} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}}): \tilde{c}_{e,m}=a_e} g_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}) \\
& = \frac{1}{M} \sum_{m \in [M]} \frac{1}{Z'_M(\mathbf{N})} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} [\tilde{c}_{e,m}=a_e] \cdot g_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}) \\
& \stackrel{(d)}{=} \sum_{\beta \in \mathcal{B}'_M} \frac{1}{M} \sum_{m \in [M]} \frac{1}{Z'_M(\mathbf{N})} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} \\
& \quad \left[\varphi_M(\tilde{\mathbf{N}}, \tilde{\mathbf{c}}) = \beta \right] \cdot [\tilde{c}_{e,m}=a_e] \cdot g_{\tilde{\mathbf{N}}}(\tilde{\mathbf{c}}) \\
& \stackrel{(e)}{=} \sum_{\beta \in \mathcal{B}'_M} \exp(-M \cdot U_B(\beta)) \cdot \frac{1}{Z'_M(\mathbf{N})} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} \\
& \quad \left[\varphi_M(\tilde{\mathbf{N}}, \tilde{\mathbf{c}}) = \beta \right] \cdot \underbrace{\frac{1}{M} \cdot \sum_{m \in [M]} [\tilde{c}_{e,m}=a_e]}_{= \beta_{e,a_e}} \\
& = \sum_{\beta \in \mathcal{B}'_M} \beta_{e,a_e} \cdot \exp(-M \cdot U_B(\beta)) \\
& \quad \cdot \frac{1}{Z'_M(\mathbf{N})} \sum_{\tilde{\mathbf{N}} \in \tilde{\mathcal{N}}_M} \sum_{\tilde{\mathbf{c}} \in \mathcal{C}(\tilde{\mathbf{N}})} \left[\varphi_M(\tilde{\mathbf{N}}, \tilde{\mathbf{c}}) = \beta \right] \\
& \stackrel{(f)}{=} \frac{|\tilde{\mathcal{N}}_M|}{Z'_M(\mathbf{N})} \cdot \sum_{\beta \in \mathcal{B}'_M} \beta_{e,a_e} \cdot \exp(-M \cdot U_B(\beta)) \cdot \bar{C}_M(\beta) \\
& \stackrel{(g)}{=} \frac{|\tilde{\mathcal{N}}_M|}{Z'_M(\mathbf{N})} \cdot \sum_{\beta \in \mathcal{B}'_M} \beta_{e,a_e} \cdot \exp(-M \cdot F_B(\beta) + o(M)).
\end{aligned}$$

where at steps (a), (b), and (c) we have used Definition 51, where at step (d) we have used Definitions 22 and 24, where at step (e) we have used Theorem 31, where at step (f) we have used Eq. (21) and Definitions 27 and 28, and where at step (g) we have used (24).

The next step is to evaluate $\eta_{e,M}(a_e)$ in the limit $M \rightarrow \infty$. Because the size of the set \mathcal{B}'_M grows polynomially in M (see Remark 26), we can use an approach similar to the one that was used in Section IV to simplify (10) in the limit $M \rightarrow \infty$. We obtain

$$\lim_{M \rightarrow \infty} \eta_{e,M}(a_e) = \gamma_e \cdot \beta_{e,a_e}^*, \quad e \in \mathcal{E}, a_e \in \mathcal{A}_e,$$

where $\gamma_e \in \mathbb{R}_{>0}$ is some suitable constant, and where

$$\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} F_B(\beta) \Big|_{T=1}.$$

Actually, $\gamma_e = 1$ because $\eta_{e,\infty}(a_e)$ was defined such that $\sum_{a_e} \eta_{e,\infty}(a_e) = 1$ for all $e \in \mathcal{E}$.

For any $f \in \mathcal{F}$ and any $a_f \in \mathcal{A}_f$, the proof of the second statement in Theorem 52 is nearly identical to the above proof. We omit the details.

APPENDIX D PROOF OF LEMMA 53

Recall the definition of the Bethe entropy function from Definition 14 and the induced Bethe entropy function from Definition 40. Fix some pseudo-codeword $\omega = \omega \cdot (1, \dots, 1) \in \mathcal{P}$, $0 \leq \omega \leq 1$, and let $\beta^* \triangleq \Psi_{\text{BME}}(\omega)$. We have to evaluate

$$\begin{aligned}
H_B(\omega) &= H_B(\beta^*) \\
&= \sum_f H_{B,f}(\beta_f^*) - \sum_{e \in \mathcal{E}_{\text{full}}} H_{B,e}(\beta_e^*) \\
&= \sum_i H_{B,i}(\beta_i^*) + \sum_j H_{B,j}(\beta_j^*) - \sum_{e \in \mathcal{E}_{\text{full}}} H_{B,e}(\beta_e^*).
\end{aligned}$$

Clearly, for every $i \in \mathcal{I}$ we have $\beta_{i,(0,\dots,0)}^* = 1 - \omega$ and $\beta_{i,(1,\dots,1)}^* = \omega$, and so

$$H_{B,i}(\beta_i^*) = h(\omega), \quad i \in \mathcal{I}. \quad (44)$$

Moreover, the edge consistency constraints of \mathcal{B} imply that for every $e \in \mathcal{E}_{\text{full}}$ it holds that $\beta_{e,0}^* = 1 - \omega$ and $\beta_{e,1}^* = \omega$, and so

$$H_{B,e}(\beta_e^*) = h(\omega), \quad e \in \mathcal{E}_{\text{full}}. \quad (45)$$

The computations for $H_{B,j}(\beta_j^*)$ are more involved because we need to find the maximizing $\beta = \beta^*$ in (35).

These computations are simplified by the observation that $H_{B,j}(\beta_j^*)$ can be maximized for every $j \in \mathcal{J}$ separately. Therefore, let us fix some $j \in \mathcal{J}$. We have to maximize

$$H_{B,j}(\beta_j) = - \sum_{\mathbf{a}_j} \beta_{j,\mathbf{a}_j} \log(\beta_{j,\mathbf{a}_j}) \quad (46)$$

under the constraints

$$\sum_{\mathbf{a}_j: a_{j,e}=1} \beta_{j,\mathbf{a}_j} = \omega, \quad e \in \mathcal{E}_j, \quad (47)$$

$$\sum_{\mathbf{a}_j} \beta_{j,\mathbf{a}_j} = 1, \quad (48)$$

where the constraints in (47) are implied by the edge consistency constraints of \mathcal{B} . (Strictly speaking, we also have to impose the inequalities $0 \leq \beta_{j,\mathbf{a}_j} \leq 1$ for all $\mathbf{a}_j \in \mathcal{A}_j$, however, we will see that the solution satisfies them automatically.) Introducing Lagrange multipliers $\{s_{j,e}\}_{e \in \mathcal{E}_j}$ and ν_j , we obtain the Lagrangian

$$\begin{aligned}
& - \sum_{\mathbf{a}_j \in \mathcal{B}_j} \beta_{j,\mathbf{a}_j} \log(\beta_{j,\mathbf{a}_j}) \\
& + \sum_{e \in \mathcal{E}_j} s_{j,e} \cdot \left(\sum_{\mathbf{a}_j: a_{j,e}=1} \beta_{j,\mathbf{a}_j} - \omega \right) + \nu_j \cdot \left(\sum_{\mathbf{a}_j} \beta_{j,\mathbf{a}_j} - 1 \right).
\end{aligned}$$

Because of the concavity of the Lagrangian in $\{\beta_{j,a_j}\}_{a_j}$, and because of the symmetry of the setup (i.e., the symmetry of the single parity-check code and the symmetry of the constraints), all Lagrange multipliers $\{s_{j,e}\}_{e \in \mathcal{E}_j}$ must take on the same value, say s_j . Therefore, the new Lagrangian is

$$\begin{aligned} & - \sum_{a_j} \beta_{j,a_j} \log(\beta_{j,a_j}) \\ & + s_j \sum_{e \in \mathcal{E}_j} \left(\sum_{a_j: a_{j,e}=1} \beta_{j,a_j} - \omega \right) + \nu_j \cdot \left(\sum_{a_j} \beta_{j,a_j} - 1 \right) \\ & \stackrel{(a)}{=} - \sum_{a_j} \beta_{j,a_j} \log(\beta_{j,a_j}) \\ & + s_j \cdot \sum_{a_j} w_H(a_j) \beta_{j,a_j} - s_j d_R \omega + \nu_j \cdot \sum_{a_j} \beta_{j,a_j} - \nu_j, \end{aligned}$$

where $w_H(a_j)$ denotes the Hamming weight of a_j , and where at step (a) we have used $|\mathcal{E}_j| = d_R$. Computing the gradient of the Lagrangian with respect to $\{\beta_{j,a_j}\}_{a_j}$, and setting it equal to the zero vector, we obtain

$$-\log(\beta_{j,a_j}^*) - 1 + s_j \cdot w_H(a_j) + \nu_j \stackrel{!}{=} 0, \quad a_j \in \mathcal{B}_j.$$

Therefore,

$$\beta_{j,a_j}^* = \frac{\exp(s_j \cdot w_H(a_j))}{\sum_{a'_j} \exp(s_j \cdot w_H(a'_j))}, \quad a_j \in \mathcal{B}_j. \quad (49)$$

We define

$$\theta_j(s_j) \triangleq \log \left(\sum_{a_j} \exp(s_j \cdot w_H(a_j)) \right). \quad (50)$$

Then the sum of all the constraints in (47) implies

$$\begin{aligned} d_R \omega &= \sum_{e \in \mathcal{E}_j} \sum_{a_j: a_{j,e}=1} \beta_{j,a_j}^* = \sum_{a_j} w_H(a_j) \cdot \beta_{j,a_j}^* \\ &\stackrel{(a)}{=} \sum_{a_j} w_H(a_j) \cdot \frac{\exp(s_j \cdot w_H(a_j))}{\sum_{a'_j} \exp(s_j \cdot w_H(a'_j))} \stackrel{(b)}{=} \frac{d}{ds_j} \theta_j(s_j), \end{aligned} \quad (51)$$

where at step (a) we have used (49), and where at step (b) we have used (50). Solving for ω we obtain

$$\omega = \omega^{(j)}(s_j) \triangleq \frac{1}{d_R} \cdot \frac{d}{ds_j} \theta_j(s_j). \quad (52)$$

With this, the entropy expression in (46) can be rewritten to read

$$\begin{aligned} H_{B,j}(\beta_j^*) &\stackrel{(a)}{=} -s_j \frac{d}{ds_j} \theta_j(s_j) + \theta_j(s_j) \\ &\stackrel{(b)}{=} -d_R \cdot s_j \cdot \omega^{(j)}(s_j) + \theta_j(s_j), \end{aligned} \quad (53)$$

where at step (a) we have used (49), (50), and (51), and where at step (b) we have used (52). It can be verified that this is indeed the maximal value of (46) under the constraints in (47)–(48).

Note that $\theta_j(s_j)$ is a strictly convex function in s_j (see, e.g., [49]), and so $\frac{d}{ds_j} \theta_j(s_j)$ is a strictly monotonically increasing function in s_j . This implies that for every $j \in \mathcal{J}$ there is a unique s_j such that $\omega = \omega^{(j)}(s_j)$.

Because all function nodes $j \in \mathcal{J}$ have the same degree, it is clear that the functions θ_j and $\omega^{(j)}$ are independent of j . This implies that there is an $s \in \mathbb{R}$ such that $s_j = s$ for all $j \in \mathcal{J}$. It also implies that $H_{B,j}(\beta_j^*)$ is independent of j .

Finally, adding up all entropy terms, the induced Bethe entropy equals

$$\begin{aligned} H_B(\omega) &= \sum_i H_{B,i}(\beta_i^*) + \sum_j H_{B,j}(\beta_j^*) - \sum_{e \in \mathcal{E}_{\text{full}}} H_{B,e}(\beta_e^*) \\ &\stackrel{(a)}{=} \sum_i h(\omega) - \sum_j d_R \cdot s \cdot \omega(s) + \sum_j \theta(s) - \sum_{e \in \mathcal{E}_{\text{full}}} h(\omega) \\ &\stackrel{(b)}{=} -n \cdot (d_L - 1) \cdot h(\omega(s)) - n \cdot d_L \cdot s \cdot \omega(s) + n \cdot \frac{d_L}{d_R} \cdot \theta(s), \end{aligned}$$

where at step (a) we have used (44), (45), and (53), and where at step (b) we have used $|\mathcal{I}| = n$, $|\mathcal{E}_{\text{full}}| = n \cdot d_L$, and $|\mathcal{J}| = |\mathcal{E}_{\text{full}}|/d_R = n \cdot d_L/d_R$.

The proof of this lemma is then concluded by observing that $\theta_j(s_j)$ in (50) can also be written as

$$\theta_j(s_j) = \log \left(\sum_{\substack{w=0 \\ w \text{ even}}}^{d_R} \binom{d_R}{w} \exp(s_j \cdot w) \right), \quad (54)$$

where we have used the fact that the local constraint code \mathcal{A}_j contains $\binom{d_R}{w}$ codewords of weight w if $w \in \{0, 1, \dots, d_R\}$ is even, and 0 codewords of weight w if $w \in \{0, 1, \dots, d_R\}$ is odd.

REFERENCES

- [1] N. Wiberg, H.-A. Loeliger, and R. Kötter, "Codes and iterative decoding on general graphs," *Europ. Trans. on Telecomm.*, vol. 6, pp. 513–525, Sep./Oct. 1995.
- [2] N. Wiberg, "Codes and decoding on general graphs," Ph.D. dissertation, Department of Electrical Engineering, Linköping University, Sweden, 1996.
- [3] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.
- [4] R. Koetter and P. O. Vontobel, "Graph covers and iterative decoding of finite-length codes," in *Proc. 3rd Intern. Symp. on Turbo Codes and Related Topics*, Brest, France, Sep. 1–5 2003, pp. 75–82.
- [5] P. O. Vontobel and R. Koetter, "Graph-cover decoding and finite-length analysis of message-passing iterative decoding of LDPC codes," *CoRR*, <http://www.arxiv.org/abs/cs.IT/0512078>, Dec. 2005.
- [6] —, "On the relationship between linear programming decoding and min-sum algorithm decoding," in *Proc. Intern. Symp. on Inf. Theory and its Applications (ISITA)*, Parma, Italy, Oct. 10–13 2004, pp. 991–996.
- [7] J. Feldman, "Decoding error-correcting codes via linear programming," Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 2003.
- [8] J. Feldman, M. J. Wainwright, and D. R. Karger, "Using linear programming to decode binary linear codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 954–972, Mar. 2005.
- [9] H. Bethe, "Statistical theory of superlattices," *Proc. Royal Society London A*, vol. 150, no. 871, pp. 552–575, Jul. 1935.
- [10] R. Peierls, "On Ising's model of ferromagnetism," *Math. Proc. Camb. Phil. Soc.*, vol. 32, no. 3, pp. 477–481, Oct. 1936.
- [11] M. Kurata, R. Kikuchi, and T. Watai, "A theory of cooperative phenomena. III. Detailed discussions of the cluster variation method," *J. Chem. Phys.*, vol. 21, no. 3, pp. 434–448, Mar. 1953.
- [12] T. P. Eggarter, "Cayley trees, the Ising problem, and the thermodynamic limit," *Phys. Rev. B*, vol. 9, no. 7, pp. 2989–2992, Apr. 1974.
- [13] M. F. Thorpe, "Bethe lattices," in *Excitations in Disordered Systems (NATO Advanced Study Institute Series B78)*, M. F. Thorpe, Ed. Plenum, New York, 1982, pp. 85–107.

- [14] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*. London, UK: Academic Press, 1982.
- [15] B. J. Frey, R. Koetter, and A. Vardy, "Signal-space characterization of iterative decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 766–781, Feb. 2001.
- [16] S. L. Fogal, R. McEliece, and J. Thorpe, "Enumerators for protograph ensembles of LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sep. 4–9 2005, pp. 2156–2160.
- [17] D. Divsalar, C. Jones, S. Dolinar, and J. Thorpe, "Protograph based LDPC codes with minimum distance linearly growing with block size," in *Proc. IEEE Global Communications Conference*, St. Louis, MO, USA, Nov. 28–Dec. 2 2005.
- [18] K. Fu and A. Anastasopoulos, "Stopping-set enumerator approximations for finite-length protograph LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, June 24–29 2007, pp. 2946–2950.
- [19] C. Ravazzi and F. Fagnani, "Spectra and minimum distances of repeat multiple-accumulate codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4905–4924, Nov. 2009.
- [20] S. Abu-Surra, D. Divsalar, and W. E. Ryan, "Enumerators for protograph-based ensembles of LDPC and generalized LDPC codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 858–886, Feb. 2011.
- [21] M. F. Flanagan, E. Paolini, M. Chiani, and M. P. C. Fossorier, "On the growth rate of the weight distribution of irregular doubly generalized LDPC codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3721–3737, Jun. 2011.
- [22] D. Divsalar and L. Dolecek, "Graph cover ensembles of non-binary protograph LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 1–6 2012.
- [23] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.
- [24] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.
- [25] D. Angluin, "Local and global properties in networks of processors," in *Proc. 12th Annual ACM Symp. on Theory of Computing*, Los Angeles, CA, 1980, pp. 82–93.
- [26] N. Ruozzi, J. Thaler, and S. Tatikonda, "Graph covers and quadratic minimization," in *Proc. 47th Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, IL, USA, Sep. 30–Oct. 2 2009, pp. 1590–1596.
- [27] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *J. Mach. Learn. Res.*, vol. 7, pp. 2031–2064, Dec. 2006.
- [28] J. M. Mooij and H. J. Kappen, "On the properties of the Bethe approximation and loopy belief propagation on binary networks," *J. Stat. Mech.: Theory and Experiment*, p. P11012, Nov. 2005.
- [29] T. Heskes, "On the uniqueness of loopy belief propagation fixed points," *Neural Computation*, vol. 16, no. 11, pp. 2379–2413, Nov. 2004.
- [30] M. Chertkov and V. Y. Chernyak, "Loop series for discrete statistical models on graphs," *J. Stat. Mech.: Theory and Experiment*, p. P06009, Jun. 2006.
- [31] J. M. Walsh, P. A. Regalia, and C. R. Johnson, Jr., "Turbo decoding as iterative constrained maximum-likelihood sequence detection," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5426–5437, Dec. 2006.
- [32] P. A. Regalia and J. M. Walsh, "Optimality and duality of the turbo decoder," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1362–1377, Jun. 2007.
- [33] J. M. Mooij and H. J. Kappen, "Sufficient conditions for convergence of the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4422–4437, Dec. 2007.
- [34] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*. Singapore: World Scientific, 1987.
- [35] M. Mézard and A. Montanari, *Information, Physics, and Computation*. New York, NY: Oxford University Press, 2009.
- [36] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [37] R. Mori, "Connection between annealed free energy and belief propagation on random factor graph ensembles," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul. 31–Aug. 5 2011, pp. 2010–2014.
- [38] R. Mori and T. Tanaka, "Central approximation in statistical physics and information theory," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 1–6 2012, pp. 1652–1656.
- [39] —, "New generalizations of the Bethe approximation via asymptotic expansion," in *Proc. 35th Symp. Inf. Theory and its Appl.*, Beppu, Oita, Japan, Dec. 11–14 2012.
- [40] H. Minc, *Permanents*. Reading, MA: Addison-Wesley, 1978.
- [41] P. O. Vontobel, "The Bethe permanent of a non-negative matrix," accepted for *IEEE Trans. Inf. Theory*, available online under <http://arxiv.org/abs/1107.4196>, Jan. 2012.
- [42] R. Smarandache, "Pseudocodewords from Bethe permanents," submitted to *IEEE Trans. Inf. Theory*, available online under <http://arxiv.org/abs/1112.4625>, Dec. 2011.
- [43] C. Greenhill, S. Janson, and A. Ruciński, "On the number of perfect matchings in random lifts," *Comb., Prob., and Comp.*, vol. 19, no. 5–6, pp. 791–817, Nov. 2010.
- [44] Y. Watanabe, "A conjecture on independent sets and graph covers," *CoRR*, available online under <http://arxiv.org/abs/1109.2445>, Oct. 2011.
- [45] N. Ruozzi, "The Bethe partition function of log-supermodular graphical models," in *Proc. Neural Inf. Proc. Sys. Conf.*, Lake Tahoe, NV, USA, Dec. 3–6 2012.
- [46] E. B. Sudderth, M. J. Wainwright, and A. S. Willsky, "Loop series and Bethe variational bounds in attractive graphical models," in *Proc. Neural Information Processing Systems Conference*, Vancouver, Canada, Dec. 3–8 2007.
- [47] F. Parvaresh and P. O. Vontobel, "Approximately counting the number of constrained arrays via the sum-product algorithm," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 1–6 2012, pp. 279–283.
- [48] P. O. Vontobel, "The Bethe approximation of the pattern maximum likelihood distribution," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 1–6 2012, pp. 2012–2016.
- [49] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [50] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [51] G. D. Forney, Jr., "Codes on graphs: normal realizations," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 520–548, Feb. 2001.
- [52] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Sig. Proc. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.
- [53] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons Inc., 1991.
- [54] G. D. Forney, Jr., "Codes on graphs: duality and MacWilliams identities," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1382–1397, Mar. 2011.
- [55] A. Al-Bashabsheh and Y. Mao, "Normal factor graphs and holographic transformations," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 752–763, Feb. 2011.
- [56] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest: Akadémiai Kiadó (Publishing House of the Hungarian Academy of Sciences), 1981.
- [57] J. Yedidia, "An idiosyncratic journey beyond mean field theory," in *Advanced Mean Field Methods, Theory and Practice*, M. Oppor and D. Saad, Eds. MIT Press, Jan. 2001, pp. 21–36.
- [58] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, Jan. 2008.
- [59] H. M. Stark and A. A. Terras, "Zeta functions of finite graphs and coverings," *Adv. in Math.*, vol. 121, no. 1, pp. 124–165, Jul. 1996.
- [60] R. Koetter, W.-C. W. Li, P. O. Vontobel, and J. L. Walker, "Characterizations of pseudo-codewords of (low-density) parity-check codes," *Adv. in Math.*, vol. 213, no. 1, pp. 205–229, Aug. 2007.
- [61] C. A. Kelley and D. Sridhara, "Pseudocodewords of Tanner graphs," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4013–4038, Nov. 2007.
- [62] N. Axiyig, D. Dreher, K. Morrison, E. Psota, L. C. Perez, and J. L. Walker, "Analysis of connections between pseudocodewords," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4099–4107, Sep. 2009.
- [63] W. S. Massey, *Algebraic Topology: an Introduction*. New York: Springer-Verlag, 1977, reprint of the 1967 edition, Graduate Texts in Mathematics, Vol. 56.
- [64] P. O. Vontobel, "Symbolwise graph-cover decoding: connecting sum-product algorithm decoding and Bethe free energy minimization," in *Proc. 46th Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, IL, USA, Sep. 23–26 2008, slides available at <http://www.pseudocodewords.info>.
- [65] —, "Counting, counting, counting (or, finite-length analysis of the sum-product algorithm)," Plenary talk at 2009 Information Theory Workshop in Taormina, Italy, Oct. 13 2009, slides available at <http://www.pseudocodewords.info>.
- [66] R. G. Gallager, *Low-Density Parity-Check Codes*. M.I.T. Press, Cambridge, MA, 1963.

- [67] C. Di, A. Montanari, and R. Urbanke, "Weight distributions of LDPC code ensembles: combinatorics meets statistical physics," in *Proc. IEEE Int. Symp. Inf. Theory*, Chicago, IL, USA, June 27–July 2 2004, p. 102.
- [68] G. D. Forney, Jr. and H. Gluesing-Luerssen, "Observability, controllability and local reducibility," to appear in *IEEE Trans. Inf. Theory*, 2012.
- [69] P. O. Vontobel, "Connecting the Bethe entropy and the edge zeta function of a cycle code," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 13–18 2010, pp. 704–708.
- [70] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, 2001.
- [71] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY: Cambridge University Press, 2008.
- [72] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: asymptotic distance distributions," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 887–908, Apr. 2002.
- [73] W. Wiegand and T. Heskes, "Fractional belief propagation," in *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, MA: MIT Press, 2003, pp. 438–445.
- [74] R. J. McEliece and M. Yildirim, "Belief propagation of partially ordered sets," in *Mathematical Systems Theory in Biology, Communication, Computation, and Finance, IMA Volumes in Math. & Appl.*, D. Gilliam and J. Rosenthal, Eds. Springer Verlag, 2003.
- [75] P. O. Vontobel, "A combinatorial characterization of the Bethe and the Kikuchi partition functions," in *Proc. Inf. Theory Appl. Workshop*, UC San Diego, La Jolla, CA, USA, Feb. 6–11 2011.
- [76] —, "A graph-dynamics interpretation of the sum-product algorithm," in *Proc. Inf. Theory Appl. Workshop*, UC San Diego, La Jolla, CA, USA, Feb. 8–13 2009, slides available at <http://www.pseudocodewords.info>.